万卷方法 | 社会科学研究方法经典译丛

抽样调查设计导论

■罗纳德·扎加　约翰尼·布莱尔 著　■沈崇麟 译

Designing Surveys:
A Guide to Decisions and Procedures



重庆大学出版社
http://www.cqup.com.cn

PDG

抽样调查设计导论

■罗纳德·扎加　约翰尼·布莱尔 著　■沈崇麟 译

Designing Surveys:
A Guide to Decisions and Procedures

重庆大学出版社

作译者简介

罗纳德·扎加(Ronald Czaja),北卡罗莱那大学社会学和人类学系副教授,为本科生和研究生讲授研究方法和医学社会学。他在方法论研究方面的兴趣主要在稀少人口总体的抽样、调查的回答效应和问卷设计中的认知问题。在从事学术工作的初期,他曾经就职于伊利诺伊州立大学芝加哥分校的调查研究实验室,先后担任过项目协调人、抽样部门的负责人、实验室主任助理和调查项目的负责人。

约翰尼·布莱尔(Johnny Blair),阿波特联合股份有限公司(Abt Associates Inc)高级调查方法师和首席科学家。他在方法学领域的试调查技术中的认知访谈、特殊或稀少人口总体的抽样设计及其他方面进行过很多研究。他曾经为不少的公共或私营部门担任过调查顾问。在进入阿波特联合股份有限公司前,在1989年到2001年,他曾经在马里兰大学的调查研究中心任职,在1999年到2001年曾经担任过中心的执行主任。刚开始进入学术界的时候,曾在伊利诺伊州立大学香槟分校的调查研究实验室任职,主要工作是抽样和调查的组织和实施。

沈崇麟 中国社会科学院社会学研究所研究员,中国社会科学院研究生院教授,"第二、三、四批百县市国情调查问卷调查"、"中国城乡居民生活调查"、"中国城乡家庭变迁调查"和"中国城乡社会变迁调查"等研究项目主持人。《社会统计学》、《社会统计方法和技术》、《实用社会科学计算机技术》以及第二、三、四批中国百县市国情调查问卷调查的《调查报告和资料汇编》、《世纪之交的城乡家庭》等译者和专著的译者或作者。

总序 社会研究方法的现状及其发展趋势

近年来,社会调查技术和社会研究方法都有很大的发展。在调查技术方面,自20世纪70年代以来,社会变迁多次横断面的跟踪调查研究,几乎成为所有国家和地区了解社会结构转变和社会发展状况的基础性调查。这种调查不仅对社会学的研究有很大促进,对整个社会科学的研究都产生了重大影响,而且这些调查结果有的已成为政府有关部门决策的重要依据。国际上比较著名的此类调查有:美国芝加哥大学全国民意调查中心(National Opinion Research Center,简称 NORC)的"社会综合调查(General Social Survey ,简称 GSS)",英国埃塞克斯大学调查中心进行的"全国家庭生活和社会变迁调查",法国经济和社会调查所进行的"全国经济社会调查",日本社会学会组织进行"全国社会分层与社会流动调查(简称 SSM)"。中国台湾"中央"研究院社会学研究所,也每两年进行一次"台湾社会变迁基本调查"。美国的"社会基础调查",现在已成为年度性的调查项目,它是美国国家基金会目前资助的最大的社会科学研究项目。以上这些调查,除美国的调查外,一般均因经费原因采用纵向的间隔性重复调查法,即每隔一段时间,进行一次全国规模的抽样调查。每次调查除保留社会研究所需的基本项目外,每次都有不同的主题。在间隔若干时间后,再重复同一主题的调查,这样的研究设计,使社会变迁研究在可以涉及更为广泛的研究领域的同时,具有更好的积累性和可比性。多年来,这些基础性调查获得的资料,滋养着大批的社会科学研究者,有时一项调查就有很多名博士生用来写博士论文,以此所取得的研究成就,其可靠性受到社会科学界的广泛认同。例如1997年出版的,以台湾社会变迁基本调查数据为基础的研究报告集《90年代的台湾社会,社会变迁基本调查研究系列二》收论文16篇,内容涉及社会生活的各个方面,在台湾引起了极大的反响。

国内社会科学界在这方面也有了长足的发展。笔者所在的中国社会科学院社会学研究所的社会调查和方法研究室,组织或参与了多项与社会变迁有关的大规模抽样调查,取得了一定的研究成果,并积累了大量有关社会变迁的宝贵数据资料,其中主要有:

1. 城乡家庭变迁系列调查:该课题是由中国社会科学院社会学研究所牵头,联合北京大学和地方社科院的研究人员展开的一项类似多次横断面的城乡家庭变迁调查。这一调查始于1981年的"中国五城市婚姻家庭调查",而后有1988年的"中国农村家庭调查"、1991年的"中国七城市家庭调查"、1998年的"中国城乡家庭变迁调查"。

2. 有关中国城乡社会变迁的系列调查:调查始于1991年的第二批国情调查,

然后有1992年的“中国城乡居民生活调查”、1993年的“第三批国情调查”、1995年的“第四批国情调查”和1997年的“中国沿海发达地区社会变迁调查”。上述调查虽然还不是严格意义上的多次横断面的纵贯研究，但研究者已在研究设计中尽量考虑到纵贯研究的基本原则，如调查队伍的稳定、指标的可比性和样本空间的延续性等。

3.中国城乡社会变迁调查：这一调查开始于2000年，为中国社会科学院重大课题。目前已经完成第一期第一次调查和第二次调查，今后将把这一调查发展为连续的、定期进行的社会变迁调查。

在纵向调查技术取得长足进步的同时，上世纪末至今，电话调查技术也有很大发展。电话调查涉及的范围几乎与个别（面对面）访谈同样全面。电话调查中使用的一系列方法，是在20世纪70年代后期和面对面调查一起发展起来的。在20世纪80年代中，电话调查开始变得很普遍，且成为许多场合中各种调查方法的首选。正如某些学者所言，一种在公共和私营部门被人们用来帮助提高决策效率的收集信息的有效方法为人们所普遍认同时，这一现象本身就具有方法论上的意义。不仅如此，电话调查还有很大的实践意义，因为它为研究者提供了更多的控制调查质量的机会。这一机会包括抽样、被调查人的选择、问卷题项的提问、计算机辅助电话访谈（CATI）和数据录入。正因为如此，今天在各种社会调查中，如果没有发现其他重要的足以放弃使用电话调查的原因，电话调查由于其独特的对调查质量进行全面监控的优点，常常成为各种调查方式的首选。由笔者翻译，重庆大学出版社出版的《电话调查方法：抽样、选择和督导》一书，也于2005年面世。

无论是纵向调查抑或电话调查，实际上都是收集研究资料的方法，而应用社会科学的发展，不仅在于调查技术，即收集资料技术的发展，还在于研究方法和分析技术的发展。近年来，无论是定性研究方法，还是定量研究方法都有了长足的发展。

首先，计算机技术的发展可谓突飞猛进，它对当今社会生活的各个方面产生了巨大的影响，在悄悄地改变着社会科学的研究风格和研究方式的同时，也大大提升了社会科学学者的研究能力。这种影响表现在研究过程的各个阶段，从理论建构（概念映射）、问卷设计（专业的问卷设计软件）、调查实施（计算机辅助访谈、计算机辅助电话访问系统、网络在线调查系统）、数据录入（光学标记识别软件）到数据分析（包括文本、声音、图像资料的处理），甚至延伸到写作发表阶段。这样的过程发生在如社会学、经济学、政治学、心理学、教育学中，促进了学科之间的相互借鉴和交叉融合，至少在研究方法上呈现这种趋势。随着计算机计算能力的大幅度提高，20世纪80年代之后，统计学领域内发生了一场“革命”，主要表现在对定类和定序变量的建模能力的大幅度提高上，以及与分布无关的统计分析模型的发展之上，特别是基于“Resampling”（包括Bootstrap、Jackknife、Monte Carlo模拟等）的建模技术*。同时，计算能力的提高还带动了基于神经网络、动态模拟、人工智能、生态进化等新兴的分析和预测模型的发展。这些进展都为定量社会科学研究提供了更

* 对于当前一些国内尚无公认译法的模型、软件等的名称，本“译丛”都只给出了英文的表述，以免造成称谓上的混乱。

多的可供选择的工具。

亚德瑞安·E·拉夫特里(Adrian E. Raftery)依据社会学家所处理的数据类型,将定量社会学在美国的发展划分为三个时代:第一代起始于上世纪40年代,交互表是其主要处理对象,研究重点是关联度和对数线性模型;第二代起始于上世纪60年代,主要处理单层次的调查数据,Lisrel类型的因果模型和事件史分析是其研究重点;第三代起始于上世纪80年代后期,开始处理诸如文本、空间、社会网络等非传统的数据类型,目前尚没有形成成熟的形态。拉夫特里的综述,虽然更强调定量社会学研究对统计学的贡献,但也大致勾勒出定量社会学在国外的发展脉络。

从分析模型的角度来看,定量分析在以下几个方向有了突破性发展:

1. 缺失值处理:由于社会生活的复杂性,社会调查数据常常出现缺失值,传统的处理方式是忽略这些缺失值,或者用均值替代。但现在则倾向于用多重插值法(multiple implation)或者其他基于模型的方法进行处理。这些技术的发展,不仅会增强我们对数据的处理能力,而且将改变我们设计问卷的方式。基于这些技术,我们在不增加被访者负担的前提下,大大增加调查问卷的内容:每个被访者只回答问卷的一部分,然后通过对缺失值的处理,获得他们对未回答部分的估值。

2. 非线性关系:线性假定是经典定量分析的一个常见假定,但在实际研究当中,线性假定只能被看作是对社会现实的一个逼近和简化。面对具体的研究数据,如果没有理论上的明确指引(不幸的是,我们常常没有中程理论的指引),我们是无法在线性模型和非线性模型之间做出取舍的。但MARS模型的出现,让我们可以从经验数据当中获得最为拟合的变量之间的函数关系,而不必预先做出线性假定。这样,理论思考和数据分析就可以实现一个互动的循环过程,定量分析就不单单是对理论和假设的简单证伪过程,而是理论思维一个重要组成部分。

3. 测量层次:20世纪六七十年代的统计模型,大多要求数据的测量层次在定距以上,如因素分析,但社会学的调查数据却大多为定类或定序数据。对应分析、Loglinear、Logit、Logistic Regression、潜类分析、Ordinal Regression、Normal Ogive Regression等统计模型的出现,大大提高了定量社会学处理定类和定序数据的能力。

4. 测量模型:基于文化、社会、心理和认知等方面的考虑,在社会学界仍有人对问卷调查在中国的效度提出质疑。抛弃"本土化"的文化执著,我们更应当关注的是问卷调查的项目反应理论(item response theory),即被访者回答问卷题器时的过程模型。这方面的进展主要表现在两个方面:一是分解测量量表的成分,如Rasch model、IRT分析、Mokken分析等,二是将测量模型与因果模型或其他分析模型结合在一起,明确把测量误差引入到分析当中,充分评估它们对分析结果的影响,如结构方程模型。

5. 潜变量模型:与测量模型相关联的另外一个发展方向是潜变量模型,例如,潜变量分层分析(latent class analysis)、潜变量结构分析(latent structure analysis)、潜变量赋值分析(latent budget analysis)等。"潜变量"这一概念表明,我们可以通过测量"显变量"来测量无法直接观察的理论概念,如权力、声望、地位等。这样,理论和现实之间,通过"潜变量"到"显变量"的映射(测量过程),就有了联接的桥梁。

6. 分析单元的层序性：在定量分析当中，我们常常强调要避免出现“生态谬误”，即分析单元的层次和结论或推论的层次不一致。与其相关的方法论争论是“宏观和微观”的问题。随着多层次模型的出现，我们可以同时考察多个层次上的问题，我们可以把个人放在其家庭背景中，再把家庭放在社区的背景下，考察个人层次的变量对社区变量的效应，或者社区层次的变量对个体行为的具体影响。在定量分析模型当中，“宏观和微观”的联接获得了建模技术上的支持。在这个领域当中，还有一个方向也值得关注：分析宏观层次的数据，对微观层次进行推论。

7. 社会网络模型：区分“关系数据”和“属性数据”，是把分析重点从个体/群体等社会单元转移到这些社会单元之间关系的第一步，社会网络模型是目前发展较快的一个定量分析领域，其理论根基是结构主义。社会网络分析目前仍然具有较浓厚的“形态学”特征（基于图论的缘故），但却为我们理解社会关系在社会空间上的形态奠定了基础，通过计算机模拟和研究社会网络的历期数据，研究社会结构的“发生学”性质模型也处在萌芽状态当中。

8. 系统动力学：如果说社会网络模型是在社会空间上拓展定量社会学的研究手段，那么，社会过程在时间上和物理空间上的属性，则是事件史模型、事件数模型、历期分析、Cox 回归、时间序列分析、Cohort 分析、状态空间模型等模型的研究对象。在这个领域，计量经济学为定量社会学研究提供了许多有益的范例。

9. 预测模型：上述模型仍然是在分析主义的范式下。有些社会学的应用研究，更强调模型的预测精度，而不是模型的认知价值，例如，社会趋势的预测。由于计算能力的提高，神经网络、基因算法、人工智能、模式识别等数据挖掘技术有了长足发展，已经出现了许多拟合经验数据的预测模型，比较成功的应用出现在计量经济学领域（如对股市的预测）。

10. 计算机模拟：对于社会学应用研究而言，研究的对象具有历史性、规模大、变迁的过程不仅漫长且表现某种渐进性，且因社会隔离/社会伦理原因无法接近或有实验禁忌等，无法直接进行观察和研究，这时计算机模拟就成为一个可供选择的替代方案。计算机模拟主要有两个类型，一是基于计算机网络的模拟：每台微机作为一个代理，整个网络作为“社会”，实时演化，如法国的 Swarm 计划；二是基于概念模型的系统，在计算机时间上，按照既定规则运行，较有名的研究是罗马俱乐部的《增长的极限》，常见的软件有 Simul，Arena 等。自然科学家对此方向似乎比社会学家更有兴趣。

定性研究方法一直是社会学研究领域中比较传统的研究方法，在社会学研究的古典时期，它甚至是社会学家手中唯一的研究方法。但随着定量研究方法在社会学研究中的广泛应用，定性研究方法就似乎越来越不受到人们的重视。但需要澄清的事实是，在定量分析模型取得飞速发展的同时，在过去的二十多年里，定性研究方法也有了长足的进步。主要表现在以下六个方面：

1. 研究素材日益扩大：除了传统的参与观察、深度访谈、专题小组访谈之外，会话、交谈、电视、广播、文档、日记、叙事、自传（autobiography）等社会过程中自然产生的素材，甚至社会学理论本身（理论的形式化），也开始进入定性分析的视野当中。所有这些资料，不仅可以以文本的格式存储，而且，新型的多媒体介质，如图像、声音和视频，作为原始的分析素材，也日益成为定性分析的新宠。

2. 分析方法更加多样:定性方法的种类在最近的二十多年中,更是有了一个质的飞跃。在比较传统的、源自语言学的方法,如内容分析、话语分析、修辞分析、语意分析、符号学、论据分析等方法之外,社会学家也创造出自己独特的定性分析方法,如施特劳斯(Strauss)等人的扎根理论、海斯(Heise)的事件结构分析、拉津(Ragin)的定性对比分析、Abbott 和 Hrycak 采用最优匹配技术的序列分析、亚贝儿(Abell)的形式叙事分析(formal narrative analysis)、鲍尔(Bauer)等人的语库建设、Attride-Stirling 等人的主题网络分析和神经网络技术应用的定性分析领域。所有这些方法的一个共同特征是,把定性研究向更加系统、更加精确、更加严格、更加形式化的方向推进。

3. 认识论基础更加多元化:现象学、释义学和本土方法论(ethnomethodology)的认识论,一直是定性分析的大本营,但近年来,实证主义也开始逐渐为定性分析所接纳,解释和阐释之间,由激烈的对立关系,逐渐演变为相互融合。

4. 研究过程更加客观规范:定性分析的一个主要问题在于阐释过程中不可避免的主观性。为了尽可能消除"解释者偏见"和主观选择性,定性分析开始遵循严格的程序模板或程序规则,并尝试引入定量分析中的"信度"、"效度"、"代表性"等概念,通过编码和对比,再加上传统的定性分析标准,如可解释性、透明性和一致性,使得定性研究的过程更加规范、阐释的结果更加客观,研究的结论更加可信。

5. 研究过程更加有效率:这主要应归功于大量计算机辅助定性数据分析(CAQDA)软件的涌现。从上个世纪 80 年代以来,定性分析过程的数字化和计算机化,已经是一个不可逆转的大趋势。这种发展趋势与定性研究者的理论取向无关,不管他们的理论立场是实证主义、符号互动论,还是本土方法论,大多数定性研究者都在自己的研究当中,开始采用计算机来辅助定性资料的分析过程。据不完全统计,目前已经有二十多种定性分析的软件,分别隶属于德国、英国、法国、美国等国家。其中,有一些软件是国外研究机构的科研成果,可以免费使用,但比较成熟的定性辅助系统大多是商业软件。这些定性分析的辅助系统,不仅使得研究者从处理大量文字材料的繁复劳动中解放出来,而且能够让研究者共享他们各自分析的细节,从而改变了定性研究的流程和研究集体之间的合作方式。同时,由于采用数据库结构,定性资料的管理也更加方便,这就为组织大型定性研究项目(包括多个研究地点、多个研究对象、历时的定性研究)提供了新的可能性。越来越多的定性研究人员开始走出他们的摇椅,坐到计算机屏幕前,湮没在访谈资料和故纸堆中的定性社会学家的形象已经一去不复返了。

6. 定性研究和定量研究的结合更加紧密:在定量分析方法的教材中,定性研究常常被看作是定量研究的前期准备工作,但定性研究者却持完全相反的观点,他们一般认为定性方法是自成一体的,可以完成从形成概念到检验假设的全部研究过程。在实际的应用研究中,定性方法和定量方法常常是交织在一起的,例如,克劳(Currall)等人在研究组织环境重要的群体过程时,通过内容分析把 5 年的参与观察资料量化,然后用统计分析来检验理论假定。格雷(Gray)和邓斯坦(Densten)在研究企业的控制能力时,利用潜变量模型把定性方法和定量方法有机结合在一起。雅各布斯(Jacobs)等人在研究比利时的家庭形态对配偶的家庭劳动分工影响时,首先用定量方法对纵向调查数据进行分析,从定量分析的结果中,又延伸出对核心

概念的定性研究。这三个研究分别代表了定量和定性方法相互融合的三个方向：①克劳等人的研究代表着定性方法的实践者试图将定性数据尽可能量化的取向，近年来涌现出的处理调查数据中开放题器的编码问题的工具软件(如 Words at, Smarttext 等，注意：它们都是由著名的统计软件公司出品的处理定性资料的软件)，处理定性资料的传统内容分析软件(如 Nvivo、MaxQDA、Kwalitan 等)也开始提供将定性资料转换到常用统计软件的数据接口，这些工具上的革新将加快这种趋势的发展。②格雷和邓斯坦的工作代表了“方法论多元论”的取向，即在应用研究过程中，通过核心概念的测量模型，把定性研究和定量研究结合在一起。③雅各布斯等人的工作则代表了一部分定量研究者对过度形式化的定量方法的不满，并试图通过定性方法加以弥补。在定量研究领域中，对“模型设定”问题的关注，是定量方法重新试图返回定性研究这种取向的另外一种表现。

与社会调查技术和社会研究方法突飞猛进的现实相比，我国学术界在这些方面的论著的出版似乎显得有些迟缓。虽然已经翻译了美国的一小部分经典定量分析教材，如布莱洛克(Blalock)和巴比(Babie)的教材，也有自己编写的一些教材，如袁方等人的《社会研究原理和方法》、卢淑华的《社会统计学》等，此外，偏重软件操作的还有郭志刚的《社会统计分析方法——spss 软件应用》、郭志刚的《logistic 回归模型——方法与应用》、阮桂海的《spss for windows 高级应用教程》等。在《社会学研究》等专业杂志上，也常常有一些定量分析的应用研究，可是专门的方法和应用模型研究却没有，也没有专门的方法研究期刊。仅就定量研究方法的介绍而言，也存在一些缺陷，主要表现在：

1. 原理和操作脱节。
2. 过分依赖某些商业软件，不全面。
3. 与中国的实证研究相脱节。
4. 不能反映当前方法研究的最新进展。

与定量研究方法相比，由于各种原因，定性研究方法的引进和介绍都比较少。在福特基金会资助的方法高级研讨班上，曾讨论过一些定性研究方法。在定性方法研究方面也有少数专著，如袁方和王汉生 1997 年出版的教程，陈向明 2000 年出版的专著。但总体说来，我们对定性研究方法还停留在初步介绍的阶段，主要的介绍也局限在定性研究的研究设计和资料收集的阶段上，对定性分析方法的介绍则没有能够反映出当代定性方法的最新进展。特别是，在定性分析工具(定性分析软件)的引进和研究上，基本上还是一个空白。虽然不乏一些出色的定性研究报告，但从方法研究上讲，我们才刚刚起步。当然，我们同时还应当注意到，在历史学领域，我国对定性资料的鉴别、考据和分析，积累了大量的经验和知识，这也应当是定性方法研究的知识来源之一，应努力加以发扬光大。

令人欣慰的是，社会研究方法的引进和出版方面相对滞后的状况终于有所改观。重庆大学出版社的编辑，以独到的学术眼光，逆当前出版界唯利是图的不良选题风气，投入了大量的人力物力，组织出版“万卷方法”。自 2004 年至今，已引进社会科学研究方法方面的专著十余种，在我国社会科学界已经引起了一定的反响。然而，更为可贵的是，重庆大学出版社并未以已经取得的成绩而自满，而是再接再

励,在原有“万卷方法”的基础上,进一步组织出版“万卷方法—社会科学研究方法经典译丛”。按我们的设想,“译丛”应该是一个开放的体系,旨在跟踪社会科学研究方法发展的前沿,引进和介绍这一方面的经典著作和最新成果。

“译丛”第一批有《抽样调查设计导论》、《社会科学研究设计原理》、《社会科学研究测量原理》、《社会科学研究分析技术》、《问卷设计手册》、《回归分析法》、《数据再分析法》、《抽样调查设计导论》、《社会网络分析法》、《广义潜变量模型》、《定性变量数据分析》和《复杂调查设计和分析方法》(书名也许有变化)等十余种,几乎囊括了研究设计、测量和分析方法的所有领域,涵盖从基础的回归分析到最前沿的潜变量分析和多水平模型等各种分析方法。无论是社会科学各专业的本科生、研究生,还是社会科学研究的学者都将从中有所收获。

“译丛”由中国社会科学院社会学所社会调查和方法研究室的多位研究人员担纲,主译者都是在社会研究方法各个领域中具有相当造诣的教师和研究人员。“译丛”的译者不仅仅把翻译看作是一个“翻译”,而且也把它看作是一次再学习和再创新。

我们期待“译丛”的出版能对社会研究方法的研究、应用和教学有所推动。

沈崇麟　夏传玲

2006年12月于中国社科院社会学所社会调查与方法研究室

译者前言

本书是根据罗纳德·扎加和约翰尼·布莱尔所著的《抽样调查设计导论》第2版译出的。作者从应用的角度,对抽样调查的设计问题进行了深入浅出的跨学科的介绍。诚如作者所言,要真正能取得调查的成功,在很大程度上取决于研究人员能否在调查设计和执行过程中的每一个阶段,将某几个为数不多的科学原理和操作指南应用于一系列关键的决策。而本书的主要目的就是向读者介绍这样一些切实可行的操作指南。它们对中小规模,乃至大规模的抽样调查都是非常有用的。

尽管本书的重点是抽样调查的设计和实施,作者对研究分析的目标问题也给予了足够的重视。强调抽样调查的设计和实施实际上是一个过程,是一个设计方案和实施逐步得到发展和完善的过程。作者认为研究者必须认真完成的工作包括:确定究竟想从目标总体了解什么样的信息;取得或制定一个包含哪一总体的抽样框;确定抽样设计方案,明确规定怎样从总体成员中抽取样本;制定估计计划,以便从样本估计值计算总体参数值;确定数据收集的方法;并对数据收集的方法步骤做出明确规定,以确保每一阶段所收集数据的质量。

作者在书中介绍了多点决策法(decision-points approach)认为对一个调查项目来讲,在存在许多导致错误的设计和实施方法的歧路的同时,却不存在什么唯一的正途。在这个意义上讲,研究设计就是,在资源有限的前提下,在若干种备择的调查方法中“因地制宜”地选出适合研究目标,误差相对较小,且切实可行的方法。

作为一本偏重应用的调查方面的专著,书中不能就抽样调查设计的许多理论和术语展开过多的讨论。因此,要真正能领会和掌握本书介绍的内容,读者也许应该在阅读本书的同时,参阅一些有关社会研究方法和社会统计学,特别是电话调查方面的专著。因为本书有关抽样调查方法的介绍,是围绕电话调查展开的。在电话调查方面,重庆大学出版社出版的《电话调查方法》一书也许是一本不错的参考书。

最后我要提醒读者的是,由于中美两国国情的不同,读者在将本书介绍的一些基本原理运用于自己的调查实践时,要根据我们的国情加以变通。

限于译者的水平,加之时间仓促,译文难免有不少错误之处,请读者诸君不吝赐教,本人不胜感激。

沈崇麟

2007年1月于北京

第2版前言

在本书第1版的前言中，我曾经说过，写作本书的目的之一是为了使读者了解，抽样调查中的许多问题至今仍然不甚了了。正因为如此，我们希望能有更多的研究者投入到调查的方法论问题的研究中，并把它作为自己的实际调查工作的一个组成部分。在本书第1版问世之后，尽管又有许多对抽样调查的具体操作方法有很大影响的新技术相继问世，但构成抽样调查的基本要素却始终未变。在对调查研究发生影响的各种新技术中，最为显著的当推互联网(World Wide Web)、手机(cellular telephones)、寻呼机(pagers)、应答机(answering machines)、来电显示(caller ID)、来电转移和拒绝接听(call forwarding and blocking)、全美不接受电话推销登记(the national Do Not Call Registry)、便携式电话号码(portable telephone numbers)、传真机(fax machines)、高性能和高速度便携式电脑(high-capacity, high-speed laptop computers)，以及个人数码秘书(personal digital assistants, PDAs)。

这些新技术对调查研究的帮助是不言而喻的，但是，它们也给调查研究造成了一定的难题和障碍。从积极的方面看，便携式电脑和互联网不仅使调查有了更多和更强大的功能，使数据的质量变得更好，而且使我们能在调查完成之后很快就能得到分析结果。但从消极的方面看，电话调查的抽样框往往都不够完整，且效率也不是很高，在拒绝回答的比率上升时，为了能与被调查人取得联系，我们还必须付出更多的努力。尽管互联网调查使我们得以在短时间内收集大量的数据，但是它的总体覆盖程度和回答率则常常是不尽人意的。互联网的覆盖程度、电话号码的概率和手机的普及所产生的各种问题都是我们所无法回避而必须面对的挑战。

无论对缺乏经验的调查研究者，还是经验丰富的调查研究者而言，新技术都对他们会有不少的帮助。互联网使资源有限的调查研究的初学者有机会既比较经济又切实可行地实施一项调查。大型组织机构，如人口普查局(Census Bureau)就曾借助个人数码秘书和无线技术的优势来提高个人数据的收集效率。但是重要的问题在于我们必须记住，尽管各种新技术应运而生，科学的抽样调查的基本原理却是始终不变的。即使采用各种各样的新技术，我们也仍然必须倍加小心地选择概率样本；也仍然必须设法使抽样框和数据收集模式能为我们提供精确的总体覆盖度。调查的问题不仅仍然必须编写精确，而且还必须明白易懂以保证答案不存在偏倚。而回答率则仍然是调查可信度的主要指标。在处理与人有关的各种问题时，道德伦理问题仍然是我们所无法回避的，实际上，在诸如隐私和数据使用这些问题上，由于用于数据联结和散发的技术的功能变得更为强大，公众对这些问题可能已经

变得更为敏感。本书这一版提出的忠告,并非是对以前提出的那些忠告的更换或改动,而是在原有基础上的改进和发展。

对于本书的这一版,以下几位人士给予了很大帮助:佩格·布兰特和凯文·斯蒂博格(Peg Brant and Kevin Stainback)对有关章节做了及时和很有帮助的评议;埃德·布莱尔(Ed Blair)对本书的初稿提出了许多宝贵的意见;布鲁斯·琦克和朱蒂·克莱恩(Bruce Cheek and Judy Cline)在文秘工作方面做了很多工作。我们愿借此机会向他们表示深切的谢意。

有两位人士和我们一起在本书这一版上工作了一年以上,他们给我们提供的帮助,其价值是难以估量的。在本书的写作过程中,伊丽莎白·伊斯曼(Elizabeth Eastman)对本书的许多草稿提出了大量有见地的专业和编辑方面的意见。我们的本科生助理阿列克斯·罗森伯格(Alix Rosenberg),在许多方面,包括文献检索、人口普查资料的查阅,对我们的各章节的草稿提出了很多宝贵的意见,并为本书做了索引。

最后,我愿向塞奇(Sage)出版社的编辑杰瑞·维斯贝(Jerry Westby)和在其他那些在最后的编辑出版工作中给我们以帮助的工作人员冯尼沙·范德拉、桑福德·罗宾逊和理查德·亚丁(Vonessa Vondera, Sanford Robinson, and Richard Adin)表示衷心的感谢。

正如我们一贯所赞赏的那样,读者如对本书有任何评论和建议,请不吝赐教,将你们的意见发到下列地址:ronc@ sa. ncsu. edu 和Johnny_Blair@ abtassoc. com,我们将不胜感激。

致谢

本书作者和青松出版社(Pine Forge Press)谨向以下各位为本书的出版做出贡献的评阅人表示深切的谢忱:

- 奥迪·L·布赖温斯(Audie L. Blevins)
怀俄明大学(University of Wyoming)

- G·戴维·约翰逊(G. David Johnson)
南阿拉巴马大学(University of South Alabama)

- 哈里·L·威尔逊(Harry L. Wilson)
罗诺克学院(Roanoke College)

- 爱德华·W·乌尔夫(Edward W. Wolfe)
佛罗里达大学(University of Florida)

- 玛格丽特·菲奇-霍瑟(Margaret Fitch-Hauser)
奥本大学(Auburn University)

- 弗雷德里克·J·克维兹(Frederick J. Kviz)
伊利诺伊大学芝加哥分校(University of Illinois at Chicago)

- 苏占尼·威尔逊(Suzanne Wilson)
伊利诺伊大学乌尔班纳—香槟分校(University of Illinois at Urbana-Champaign)

- 丽塔·科克·维罗克(Rita Kirk Whillock)
南卫理公会大学(Southern Methodist University)

目录

第1章 导 论

An Introduction to Surveys and to This Book

我们很难找到一种社会科学研究方法,其发展速度能与抽样调查比肩。我们也无法找到任何一种其他的社会科学研究方法,它能像抽样调查那样如此普遍地深入到我们社会生活中的各个方面。在不到两代人的时间内,采用比较小的概率样本的方法测量各种社会现象,尤其是测量公众态度和行为有了很大的发展。这种方法几乎无处不在,它涉足的领域小到调查不太引人注目的好奇心,大到调查重大的社会实践。为了了解抽样调查的发展,我们首先来关注一下抽样调查所涉足的重大的全国性问题,诸如大选前选民对各派候选人所持的态度、经济问题中的消费信心、失业的严重程度、十年一度的人口普查的精确性、各种不同类型的犯罪行为的发案率,或公众对诸如有可能感染免疫缺陷综合症(艾滋病)的危险行为的了解程度等。在市场领域,抽样调查被用于在新产品引进之前测试公众的反应,以及公众对各种商品和服务的满意程度。在学术界,抽样调查在经济学、社会学和心理学等诸多领域为社会科学家提供测试模型和检验假设的数据。实际上,正因为抽样调查方法已经在众多的研究和公共政策领域取得了巨大的成功,所以我们更有必要加倍小心,以免将抽样调查误用于不适合使用它的领域。

这门我们寄予厚望的科学,如果还不能说至今仍处在孩提时期的话,至多也不过是一个青少年。在某些方面它已经比较成熟强壮,而在某些方面则仍然缺乏经验和不够可靠。在它涉足的诸多领域中的某些领域,抽样调查对功能强大的数学工具的依赖颇多。但在一些基础理论尚未完善的领域,它却依赖于主观经验。而还有一些领域,它使用的仅仅是那些最近才进入专业人士的思考范围的规则和惯例。

如果抽样调查中存在的各种各样的问题会使躬耕于这一年轻学科的专业的实际工作者感到困惑,那么它便有可能令那些需要实施一个调查,或对调查结果做出解释的初涉调查业的新手或临时性的用户,因心存畏惧而裹足不前。

调查研究方法总论

调查研究天然具有跨学科的特点。抽样和估计方法需要概率理论和统计学知识。数据收集者必须首先说服被调查人，然而，这在某种程度上需要调查员和被调查人之间的互动。访谈和问卷则有赖于认知、会议、语言表达能力和论述。便携式电脑的使用和在互联网上进行的问卷调查则需要用各种软件进行编程。调查研究涉及诸多的学科，如统计理论、语言心理学或社会语言学等。这些学科中的每一个都关乎调查研究的某一方面，而同时通晓这些学科的有实践经验的调查研究专家却是凤毛麟角。每当在这些已有的学科中再加入一种新的学科时，那么即使是实施一个中等规模的抽样调查，必要的管理手段和专门技术也是不可缺少的。调查业似乎的确有些令人望而生畏。然而，规模非常大和非常复杂的调查始终都在进行着，尽管在某种程度上我们可以确定，它们通常可以产生有用、可信和有效的结果。但在有些情况下，它们可能无法做到这一点。

调查企业之所以能在调查中取得成功，很大程度上都应该归功于他们的研究人员能在调查设计和执行过程中的每一个阶段，将某几个为数不多的科学原理和操作指南应用于一系列关键的决策。本书将要向读者介绍的内容之一便是诸如这样的操作指南。尽管这样的介绍不能产生上面所提到的那样高水平的专家，但是它却能为读者提供切实可行的实际指导，这些指导对一个小规模或中等规模的抽样调查将是非常有用的。

即使是中等规模的调查，一般也需要大量的时间、物质材料、资金和各种帮助。当然我们的确可能发现只有一个研究者做的调查，但这种情况是一种例外。与某些科学或学术性的工作不同，抽样调查通常是一种许多有着各种技能的人参与的团队工作。即使那个提出研究问题的研究者本人能在提出问题的同时也承担问卷设计和数据分析的工作，他也仍然需要有人来帮助他收集数据、将数据输入计算机文件和完成其他各项重要的工作。因此，无论一个调查是一个企业的业务，还是一个课堂作业，都必须要有一定的分工、必要的协调、管理以及一定的时间和资金。而要在完成所有这些工作的同时，为了最大限度地提高调查的质量，研究者还必须做出很多决策。

一个以如此明智的一系列决策为指南的调查项目，将会最为充分地利用它的所有资源——不仅包括资金、物质材料和人力，而且还包括时间、信息和智力。而正是在这些决策问题上，研究者经常需要有所折中或权衡，它们正是本书所要阐述的重点。我们将要向读者阐述的是，在研究的问题确定之后，研究者怎样才能将预定的思想贯穿在调查的各个方面和问卷设计、抽样及数据

收集等各个阶段。为了更为具体地说明这一过程,我们精选了一些设计和执行互补的实际性研究项目。本书的每一章,将对这些项目的决策问题做重点介绍,与此同时我们也将引进一些其他的例子,对这些问题做进一步的介绍。

抽样调查并非适用于所有的场合。因此在开始讨论抽样调查一般原理和方法之前,我们首先将向读者介绍几条指导方针,供读者在确定调查适合某些场合时进行参考。总之,如果我们将调查运用于不恰当的场合,那么我们便几乎不可能有成功的希望。

所谓抽样调查是从一个定义清楚的总体所抽取的样本中收集信息(通常使用问卷)的方法。问卷,有时也称作工具(instrument),通常包含一系列要求被调查人回答的相关的问题。在大多数情况下(但不是在所有的情况下),问题的形式都是封闭式的,即被调查人回答问题的一组答案是事先规定的。调查得到的数字型或定量型数据将在随后被输入数据文件,以备统计分析使用。

有关调查的这一简单描述,含有几个简单却是基本的可以判断采用调查方法是否恰当的条件。要实施一个调查,首先目标总体必须定义清楚。在一般情况下,它都是人口学特征和地理学边界的简单组合。例如,一个目标总体可以是所有在马里兰州的有电话的住户中年龄在 18 岁及以上人口。所谓总体的定义清楚是指其成员的定义必须是明白无遗的。

其次,我们必须确信,大多数被调查人对我们要求他们提供的信息是有所了解的。问被调查人诸如他们的家庭财产的净值是多少这样的问题是没有什么意义的,因为他们中的大多数人,可能都无法回答。

最后,分析的目的必须是回答研究提出的问题、检验假设、估计总体特征值、为一组变量建模,或其他借助恰当的统计方法达到的定义明确的目标。

抽样调查的用途

使用复杂的统计方法,如多元或概率比对数(logistic)回归、层级分析(hierarchical analysis)和方差分析,研究者可以用调查数据进行检验假设及研究变量之间的因果关系。调查数据和与之匹配的高级分析方法,与理论构建和实验室中的心理实验一起,构成了社会调查中最前沿的工具之一。但是就最为基本的意义而言,调查发端于我们了解总体中那些未知的特征的愿望,即我们所说的测量。那样的特征通常都是一个颇为简单的参数,如美国成年人口中抽烟人口或拥有 DVD 播放机人口的比例、雇佣家庭帮佣(domestic help)的总数、能上网的住户的比例和特征,或者年平均娱乐支出(不过我们必须对娱乐支出加以定义)。当然更为复杂的统计量,例如储蓄或投资与收入的比率也是可以测量的。调查也可以只涉及总体的某一部分,如老人(年龄在 70 岁

或者以上的人)、有小孩(年龄在5岁以下)的住户,或拥有计算机的人。研究者可以在关注总体总人口的结果之外,同时也对总体中的某一或某些子群体感兴趣,这样的情形也是经常可以遇到的。例如我们可能在关心总体所有人口对犯罪问题的看法的同时,也对不同种族的人口对犯罪问题的看法感兴趣,或对城市居民与市郊居民的态度差异感兴趣。另一方面,我们也可能对总体的情况根本就没有兴趣,而只是对农民对杀虫剂对环境的影响的看法或大学生对每学年的平均费用的看法感兴趣。

许多有关调查的方法论,特别是抽样理论,其目的在于以一种比较严格的方式求得诸如总体参数这样的估计值。本书关注的重点大多数与此相同。然而,我们必须了解,这些抽样调查方法也同样适用于模型构建和总体描述。这就是说,不存在什么只适用于假设检验的抽样调查理论,当然也不存在什么只适用于政策分析人员的、用以测量诸如上面提到的那些问题的抽样调查理论。实际上,抽样调查的出众之处正是它得到的数据既可以用于描述,也可用于分析和推论。例如,政府主持的调查使我们得以了解在有正当住所的人中,需要寻找医疗保健的人的比例是多少(全美医疗卫生调查),或上个月成年人中失业人口的总数是多少(当前人口调查)。社会学家、经济学家和其他的研究人员或实际工作者,都在使用这些同样的数据,而他们的兴趣并不在得到这些简单的点估计值。

调查过程揽要

尽管本书的重点是抽样调查的设计和实施,但是我们的讨论必须从怎样明确研究分析目标开始。我们应该认识到抽样调查的设计和实施实际上是一个过程。在这一过程中,抽样调查的设计和实施逐步得到发展和完善,而非从一开始就是完备的。我们应该在一开始就有一个分析计划。这一计划应该回答某些问题,如我们的目的是否只是得到某些点估计值,还是与此同时也希望得到一些更为复杂的量度值?哪些变量对我们的研究至关重要?哪些则比较次要?我们是否要对特定的子群体进行估计?我们究竟要做一些什么样的分析?我们的数据表将会是什么样的?如果我们的目标有多个,那么其中哪些更优先?我们想要回答的研究问题究竟是什么?这些问题的答案将主导我们的调查设计。分析计划中的某些部分也可能在开始的时候无法确定,或在问卷设计的过程中有所变化。在工具的设计和改进期间,我们考虑的问题都非常具体,例如我们如何使用特殊问题,它们的确切含义又是什么等。一份问卷的长度可能超出了我们的忍受极限,这样我们就必须决定哪些问题是应该保留的,哪些问题是应该舍弃的。正因为如此,为了制定一个非常明确的分析计

划，我们不仅必须在开始制定计划的时候费很多的口舌，而且在整个研究过程中也仍然如此，因为现实和我们的想象是很不同的。

不论以后问卷怎么使用，研究者都必须认真地完成以下各项工作：确定与我们需要获取信息有关的目标总体；取得或制定一个包含那一总体的抽样框；确定抽样设计方案，明确规定怎样从总体成员中抽取样本；制定估计计划[1]，以便从样本估计值中计算总体参数值；确定数据收集的方法；对数据收集的方法步骤做出明确规定，以确保数据收集的每一阶段的数据质量。

这些步骤中的每一步都关系到如何在各种备择方法中进行选择的问题。而这些技术性的决策又不可避免地与资源的消耗问题有关。本书将在资源固定的前提下，为怎样进行决策的问题提供一些指导。这些资源有时与财政有关，但在更多的时候（尤其是对那些和志愿者一起工作的新手而言），调查必须在有足够的时间或得到其他非财政性支持的情况下进行。与大规模的政府或商业调查不同，我们的调查模式是费用固定的模式。这些调查的目的是在最大限度降低费用的同时，使调查数据达到一定的精确度。这就是说，我们的目的是在规定的"预算"之内，得到最好的测量数据。

这种所谓的多点决策法（decision-points approach）隐含着这样一种观点：对一个调查项目来讲，在存在许多导致错误的设计和实施方法的歧路的同时，却不存在什么唯一的正途。相反，可能有几种不同的方法适用于某些特定的研究目标、目标总体、资源和其他限制条件。（对以上提及的那些类型的目标）我们采取概率抽样，赋予每一个总体单位一个预知的（非零的）被抽中的概率，这通常比采取便利抽样（又名就近抽样，译者注），即在最方便取得信息的人群中进行抽样调查的效果要好。然而，可供我们选择的概率样本设计方案却有很多，因此如何选择的问题是很重要的。不仅如此，每一种设计方案都需要做出一系列与调查实施方法有关的进一步的选择。

我们试图编写的是这样一种问卷，它不仅能清楚地表达我们所需的信息是什么，而且还能使被调查人在对它的理解上不存歧义。这样的任务并不是想象的那么简单，因为与样本设计相比，在这方面可以遵循的科学原理不是很多。我们都知道我们自己在想什么，或知道希望通过问题表达什么，而这常常成为了他人不能很快理解我们的意图的障碍。一旦我们开始测试问卷的初稿，问题中存在的我们意想不到的含义立即就会出现。测试问卷的方法有很多，我们将会对这些问题展开详细的讨论。不过我们在这里要指出的是，测试能给我们的帮助固然很多，但其中很大一部分似乎来自那些只是看过我们的问卷而并未参与问卷起草的人们。这些人既可能是潜在的被调查人、朋友或同事，也可能是调查方面的专家。因为他们没有参与问卷的起草，所以并不了解问题的确切意图，只能根据它的字面和上下文来推断它的含义。

研究者可能很容易低估上下文对单个问题的影响，但是，当我们把一个个

单独的问题组合起来编成一种测量工具的时候,每一个问题则都有可能对它后面的那些问题的解释产生一定的影响。诚如卡隆(Caron,1992, p. 150)所言:"所有的语言都不可避免地存在模糊性,这种模糊性充斥于言辞之中,在构建和表达含义时具有决定性的作用。"尽管如此,我们为避免文字的模糊性所做的种种努力既不能过分简化,也不能损害一个问题的本意。

在收集数据的时候,我们不仅要考虑到我们所选择的方法是否适合我们将要提出的问题的类型,而且还要考虑到它们是否可能得到被调查人的合作。但是这些考虑本身并不会告诉我们,邮寄式调查、互联网调查、电话调查、集体调查和个别访谈,或者它们的各种组合究竟哪一种最适合我们的研究。就算调查方法选定了,它也不会引导我们顺利地通过各种前后相继的步骤,完成整个数据收集工作。本书将在诸如这样的基本决策和方法策略上给读者提供一些指导。

从始至终,我们同样关注那些有替代意义的选择的效果。例如:是否通过减少清单(抽样框)的覆盖面以降低调查方案的费用、是从数目较小的被调查人那里搜集更多的数据还是从数目更大的被调查人那里收集较少的数据、是选择较大的样本还是较高的合作率、是排除敏感问题还是准备冒被调查人拒绝回答的风险等。与此同时,结果问题和数据收集过程的质量控制问题也必须引起我们足够的注意,它们都需要耗费一定的资源。

构成我们的方法的最后一个原则是苏德曼(Sudman,1976)的一个忠告。苏德曼说,我们必须牢记,一个调查究竟需要好到什么程度,完全视我们要达到的目的而定。当我们作为调查研究者在为使一个高质量的项目变得切实可行的时候,我们也必须了解,并非所有的调查都需要达到相同的精度和信度。一个高校教员在得到某种资助后所进行的调查项目,通常都会对调查质量有一定的要求,但是本科生做的课堂调查项目可能就没有。而二者都不需要达到重要的政府项目,如当前人口调查(the Current Population Survey,CPS)或全美健康卫生调查(the National Health Interview Survey,NHIS)所要求的那种精度。正因为如此,我们在力求使每一步都能做得最好的时候,必须采取比较实际的态度,既要考虑实际的需要,也要考虑实际的可行性。这样一种认识将会在许多需要进行选择的时候给我们以帮助,指明选择的方向——不仅如此,它甚至能帮助我们确定调查是否应该继续进行下去。然而,要能真正达到这样一种认识并不是一件容易的事。

一旦我们已经确信调查方法不仅适合我们的研究目的,而且是切实可行的,我们就应立即开始确定调查的主要需求。首先,我们必须确定时间框架和预算。如果研究是1~2个学期的教学计划的一部分,那么在什么时间数据应该可以使用?如果研究是为校外的客户进行的,那么调查的时间必须根据客户的要求而定。在有关项目的日程已经确定之后,我们立即就会面临一系列

决策的问题,包括必需的样本容量和每个访谈需要多少时间。在开始制定具体的调查设计之前,我们至少对研究开展的条件有一个大致的了解,并确定它们是否都在我们掌握的资源范围之内。

不要过早地开列资源清单,应该先将我们所掌握的资源与我们的调查目的进行一番比较。进而继续以这样一种方式,为我们的调查项目制定一个比较宽泛的计划,并准备好为我们遇到的第一个重大的选择问题——数据收集方法的选择做出决策。

本书摘要

在开始讨论数据收集问题之前,我们将首先在第 2 章向读者介绍一下设计和完成一个调查所需的基本步骤:调查设计和初步计划;决定抽样方法;试调查;终结调查设计和计划(final survey design and planning);数据收集;数据编码;数据文件构建和数据分析。该章强调调查过程的关键问题和许多必须做出的决定。我们以一个有关全国电话调查的时间进度表作为该章的结尾。在时间进度表中,我们列出了调查的主要任务和完成每一项任务的时间安排。

第 3 章我们对邮寄式调查、互联网调查、电话调查和面对面调查做了简单的介绍,并就每一种调查方法需要的资源、问卷设计和数据质量,指出了它们的长处和不足。我们就费用、收集数据需要的时间、对问卷长度和复杂性的限制、可以提出的问题的类型、可视性辅助手段的使用、与被调查人建立融洽关系的可能性和回答率的高低等方面,对各种数据收集方法进行了比较。为了阐述得形象直观,我们除使用了几个来自实际研究的例子之外,还使用了几个若干种调查方法组合使用的例子。

在第 4 章,我们对问卷设计问题做了介绍。我们首先讨论了问卷设计的具体过程。这种从过程着手讨论问卷的方法既为我们提供了讨论的框架,又为新入此门的研究者提供了可以照猫画虎、亦步亦趋的步骤清单。接下来是一份优质的问卷所应具备的一组简单特质,然后,通过几个例子介绍了构建单个问题的方法,重点是如何在我们希望收集的信息和被调查人能提供的有效和可靠的答案的能力二者之间保持平衡。我们也为能把什么样的问题收入测量工具提供了指南。在第 5 章继续讨论问卷的设计问题,重点是普通问卷的设计。在我们看来,问卷的开头至关重要,我们给它的关注甚至多于前面提到的且正越来越引起关注的调查响应和可能出现的严重的无回答偏倚问题。在这一章,我们设法使读者了解如何组织和撰写问卷开首的引导语、中间的各个章节和结尾。第 6 章主要介绍如何对问卷进行综合性的测试。我们吸收了近年来在工具设计方面出现的大多数最新成果,并给读者一一介绍了主要测试

方法及如何对测试结果进行解释。在这一章我们还扩展了一节,专门用来介绍行为编码(behavior coding),更新了对有关认知访谈的讨论,以容纳那些近年来出现的对试调查问题很有意义的研究发现。

在第7章,我们介绍了抽样基础理论。首先,我们指出了非概率样本和概率样本之间存在的差异。本章的重点是介绍概率样本,并形象直观地说明它为什么是各种抽样方法的首选。对那些在设计和选取概率样本中举足轻重的关键问题,我们予以了特别的关注。这些关键问题有:定义总体、构建和评估样本清单及如何处理意想不到的和常见的问题。在这一章,我们也对经常出现的"调查的样本容量究竟需要多大"这一个问题做出了回答。假设检验和效力问题也在这一章进行了讨论。我们也用了一些例子来说明,在制定调查计划和估计为达到预期分析精度与重要的子样本群体所进行访谈的次数是否足够时,如何利用普查数据。

第8章给读者提供了若干抽样方面的实例。我们介绍了许多以电话簿为依据的社区电话调查所必须做出的决定,以及如何选取以清单作为辅助手段的全国性随机号码电话调查的样本。我们对如何在调查样本户内随机选取回答人的方法步骤进行了讨论。本章的结尾是两个如何从一些清单中选取学生的例子。第9章的内容与如何对数据收集过程中的非抽样误差进行综合处理有关。本章的前半部分介绍来自数据收集,尤其是来自单位和题项的无回答的偏倚和误差的基本性质。这一部分的结尾概括介绍了调查质量的主要量度。本章后半部分的重点是介绍降低访谈式调查和自填式调查的非抽样误差的方法。

第10章关注的问题是调查的道德伦理问题、调查预算编制及如何准备和撰写有关方法论的报告。这样一个对于精心操作的调查十分重要的问题,却常常为入门性的教科书所忽略。我们将向读者说明,完整地介绍调查采用的方法是多么重要。调查质量的各个方面都做了明确的定义,它们的主要组成部分也都做了详尽的描述。然后,我们给读者列举了许多例子,告诉读者报告中应该包括哪些内容和应该怎么样来表达这些内容。本章结尾,我们给了读者一个指南,以帮助读者了解调查的理论和方法在近年来出现的重要文献。

第2章 抽样调查的步骤

Stages of a Survey

图2.1中展示了设计和完成一个抽样调查所需的五个一般步骤，如下所列：

1. 调查设计和初步计划。
2. 试调查。
3. 终结调查设计和计划。
4. 数据收集。
5. 数据编码、数据文件构建、数据分析和最终研究报告。

本章将对每一步所需做出的决策进行讨论。下面所列的内容是设计和执行一个抽样调查的关键所在（见图2.1中的粗体方框）：

- 设计抽样调查：做出数据收集和抽样方法的初步决定。
- 试调查：对若干根据实践经验做出的决策进行检验和评估。
- 修订调查设计和操作计划：根据试调查的结果确定最终的设计方案。
- 收集数据：进行数据收集，并执行质量控制方案。

步骤一：调查设计和初步计划

步骤一涉及某些在第1章中所讨论过的问题：详细阐述我们研究的目的和问题及抽样调查将要探讨处理的研究问题。在这一阶段，我们必须确定研究的目的和怎样在时间和资源有限的前提下最好地达到这一目标。

- 研究的目的是否是检验某一假设？例如：

女性是否比男性更加认同不管是什么原因都应该允许一个妇女流产这样的观点。

- 研究的目的是否在于检验一系列相关假设组成的因果模型？例如：

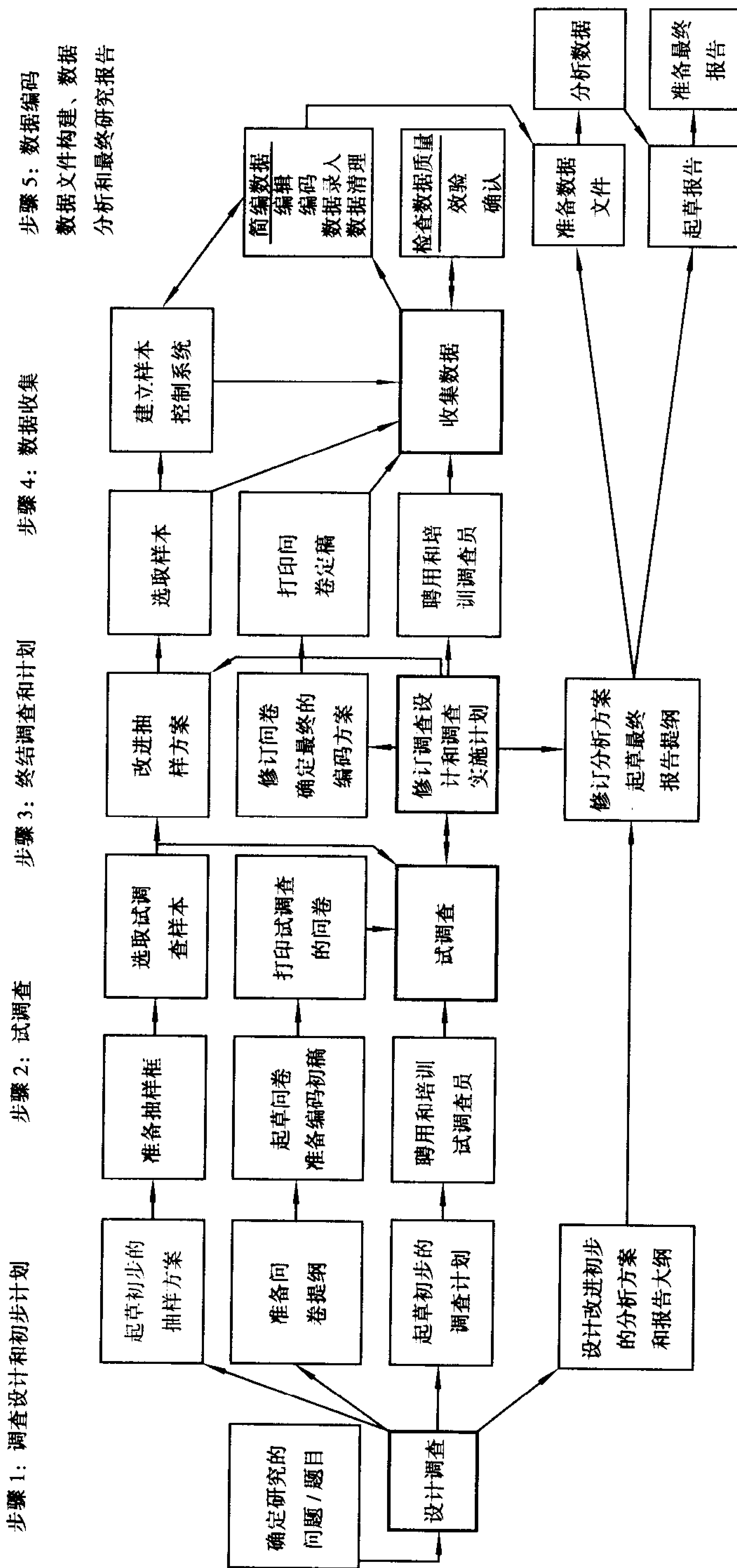

步骤1：调查设计和初步计划
步骤2：试调查
步骤3：终结调查和计划
步骤4：数据收集
步骤5：数据编码
数据文件构建、数据
分析和最终研究报告
确定研究的问题/题目
设计调查
起草初步的抽样方案
准备问卷提纲
起草初步的调查计划
设计改进初步的分析方案和报告大纲
准备抽样框
起草问卷
准备编码初稿
聘用和培训试调查员
选取试调查样本
打印试调查的问卷
试调查
改进抽样方案
修订问卷
确定最终的编码方案
修订调查设计和调查实施计划
修订分析方案
起草最终报告提纲
选取样本
打印问卷定稿
聘用和培训调查员
建立样本控制系统
收集数据
简编数据
编辑
编码
数据录入
数据清理
检查数据质量
效验
确认
准备数据文件
起草报告
分析数据
准备最终报告

图 2.1 抽样调查的步骤

1. 具有传统价值观的人使用安非他命(amphetamines)的可能性较小。
2. 有吸毒的朋友的人更有可能使用安非他命。
3. 具有传统价值观的人有吸毒的朋友的可能性较小。

- 研究的目的是否在于估计具有某种信仰或具有某种特定行为的人的比例? 例如:

总体中认为我们的刑事犯罪司法系统运行情况良好的人的比例是多少?
总体中在去年犯有抢劫罪的人的比例是多少?

- 调查的目的是为了研究某些特定的问题或人口中某一群体的态度、行为随时间的变化而变化的情况,还是为了解我们是否能找到一些干预手段来纠正人们的某些行为? 例如:

社会综合调查(The General Social Survey)在过去的三十年间,进行过很多方面问题的调查,如流产、宗教设施、人们对死刑的态度、社会等级的自我评价、邻里地区的安全问题等。

全美青年纵向调查(The National Longitudinal Survey of Youth)从 1979 年起追踪调查 12 000 个以上的青年人。这些青年每隔 1 或 2 年就要被调查一次,调查的内容包括从事过的工作份数、每次工作的收入和其他与劳动力市场有关的问题。

戒烟的社区干预实验(The Community Intervention Trial for Smoking Cessation)便是一项历时 4 年、涉及 22 个社区的干预实验。其目的在于帮助吸烟者,特别是烟瘾很重的吸烟者(每天吸烟等于或多于 25 支)戒烟,并能坚持不再抽烟。

在上述各种不同的调查目的中,有许多是典型地需要对成年人口总体的常规调查研究,而有些则可能与一般人口的调查研究无关,只是与某些特定的子群体有关。例如我们可能的调查目的只是想了解诸如牙科医务工作者对国家颁发的从医许可证条件的变更有什么想法、大学的教职员对加入工会组织持什么态度、外来移民工花费在看病上的钱占他们收入的比例是多少这样的问题。我们的研究目的也可能与个人完全不相干,但却与组织、群体、企业或政府机构有关。例如我们可能想了解某一类型组织的会员年费、每一所高中或州内每一校区高中的低年级和高年级学生的辍学率。调查研究目的的确定将为下面的一系列决策提供框架:什么人或什么事情是我们感兴趣的总体? 我们应该采用什么样的方法来收集数据?

就有关抽样的问题做出决策

在第一阶段,我们必须回答的有关抽样的基本问题是:什么人或什么事情是我们感兴趣的总体? 在对个体进行研究时,我们必须考虑的基本问题有三

个:合格的被调查人的年龄范围;研究结果想要概括的地理区域;总体包括的是那些住在家里的人,住在集体宿舍的人,还是无家可归者,或这三类人的某些组合。例如如果我们要对一般的人口总体进行抽样调查,那么我们的总体中应当包含18岁和18岁以上的人,还是应当包含21岁和21岁以上的人,或17岁以上的人?在进行诸如这样的决策的时候,我们必须以研究目的为根据。如果我们进行的是一项有关公众对公共交通工具上犯罪问题的感觉方面的调查,那么我们的被调查人就应该既包括那些使用公共交通工具的人,也应该包括那些不使用公共交通工具的人。我们应当以那些有规律地使用公共交通系统的人的年龄结构来设置合格的被调查人的年龄下限。如果我们调查的地区,学生都使用公共交通工具上学,那么年龄的下限也许应该设在16岁,甚至14岁。

在研究对象是某些群体时,我们必须确定谁是群体或组织中能为我们提供精确信息的最为关键的人物。例如,如果我们的兴趣在于了解某种类型的外科手术的平均费用和住院时间,那么我们就应该去调查我们样本中的每所医院的财务主管。有时,为了确定谁是最为关键的人物,我们需要打电话到调查地或组织的所在地。我们也可以与组织的领导联系,要求他们或某些知情者接受调查或填写问卷。如果我们需要从每一个群体了解各种不同类型的信息(如白内障手术的一般费用、生育数、就诊的癌症病人数),那么我们可能从一个组织中的两三个不同的个人那里搜集信息。在调查了解一个群体的情况时,最重要的问题是确定或找到知多识广的被调查人——也被称作知情者,这也正是这一类调查的关键之所在。

另一个必须考虑的问题是调查的地理区域是否有代表性。我们继续来讨论,公众对公共交通工具上犯罪问题的感觉方面的调查。在有些城市里,公共交通系统不仅在市中心运行,而且也在某些郊区运行。然而,郊区公共交通系统的覆盖面,不仅一般都不如市中心那么广,而且通常由不同公司提供。我们必须确定我们的调查包含的面究竟应该有多广?我们关心的公共交通系统问题究竟应该仅限于市中心,还是整个地区?或市中心和某几个郊区?在做这样的决策的时候,不存在什么魔术般的公式或经验法则。我们所必须或只能依据的东西是清晰的研究框架和可资利用的有关资源。

我们在做这些决定时,必定会有所取舍。在确定提高或降低合格调查人的年龄限制时,参考一下人口普查的资料将对我们的决策不无帮助。正如我们在后面将要讨论的那样,在其他条件相同的情况下,在某一给定的群体或子群体中,被调查的人越多,调查结果的置信度就越高。例如,假定我们电话调查能够调查的人数是500个,如果我们将合格被调查人定义为21岁及21岁以上的成年人,那么所有500个调查将都在这一年龄组的人中进行。然而,如果我们将合格的被调查人的年龄扩大到14岁及14岁以上,那么我们要调查

的年龄在 21 岁及以上的人数必定会因此而低于 500。如果我们从人口普查的数据得悉 14 ~ 20 岁的人占合格的被调查人口的 13%，且我们再假定调查合乎规范，那么大约就会有 65 个（500 的 13%）年龄在 14 ~ 20 岁的被调查人接受了调查，它将低于年龄在 21 岁及以上的接受调查的人数。这样，在我们整个调查结果的置信度并未发生变化的同时，从年龄在 21 岁及以上的被调查人中得到的结果的置信度却发生了变化，因为来自这一年龄组的被调查人的人数较原来要少。在考虑研究区域是应该包括还是排除郊区的被调查人这一问题时，我们也必须做出与此类似的取舍。而取舍应遵循的逻辑则完全相同。如果我们的预算是固定的，且决定包括部分或全部郊区，那么原来分配给城市的被调查人，就将因此而分配给郊区，城市的调查结果的置信度就会因此而有所降低（有关如何进行取舍的问题我们将在第 7 章进一步讨论）。

在组织、机构或企业研究中，我们也必须考虑类似的问题。例如，假如我们的兴趣是研究学校的体制问题，那么我们就必须确定我们的研究对象是否包括公立学校和私立学校。如果我们感兴趣的总体是医院，那么我们就必须确定我们的兴趣只是在盈利的医院，还是除此之外，对非盈利的医院也感兴趣。此外，我们可能只希望研究那些病床数最少或提供特殊服务的医院。解决这些问题的办法是周密全面地考虑我们的研究目的和估量我们手中可资利用的资源。而诸如这样的决策，是我们作为一个调查的负责人所无法回避的。一般讲，对这些我们必须做出的决策而言，不存在绝对正确或错误的答案。我们必须比照我们手中所掌握的资源和时间进度表，评定哪些决策应该更优先，并全面地衡量它们的得失。

关于抽样我们必须考虑的问题是据以抽取随机样本的抽样框的有效性。因为我们正是从抽样框中抽取研究总体的随机样本。抽样框是一种原始资料（或若干种原始资料），其中包括了合格的研究对象或群体。有时抽样框与我们的研究总体完全匹配。例如，一张清单或一份原始资料，其中包括了我们感兴趣的总体的每一个个体或学校。而在另外一些时候，抽样框所包含的内容可能比我们感兴趣的总体更多，因此我们就需要对样本成员进行筛选，以确定他们是否合格。还有一些时候，抽样框并没有包括全部的合格的总体成员。在遇有这样的情况时，我们必须确定，就我们的研究目的而言，这样的抽样框是否足够精确。下面是两个有关抽样框的实例：

1. 如果我们想要对一所大学所有注册的学生进行调查，那么我们就可以到注册办公室去要一张登有所有目前在册学生的清单。除了名字之外，清单上还记录了他们的地址和电话号码。这样的清单就可以作为我们的抽样框。

2. 如果我们想对北卡罗莱纳的成年人进行一次电话调查，那么我们可以设法找到这么几本电话簿：只要它们合在一起，便可得到该州所有地区的电话号码。这样，这几本电话簿便可作为我们这次电话调查的抽样框[1]。

就各种一般人口总体的调查而言,电话簿是诸多使用最为普遍的抽样框之一。不过,在用电话簿作为抽样框时,还会出现一个比较重要的问题:就合格的人口总体而言,电话簿的代表性究竟如何?什么样的人可能没有被包含在电话簿内?不言而喻,那些没有电话的人,以及那些未登录电话号码和大多数只有手机的人是不可能包含在里面的。

抽样框在很多情况下都未能包括合格总体中的所有的个体。在总体定义与抽样框不相匹配时,我们需要考虑以下问题:丢失的总体究竟有多少?丢失的究竟是一些什么人?如果将他们排除在外,那么对因变量,即我们希望解释或估计的变量的结果的影响究竟有多大?实际上,我们要考虑的问题是:我们的调查允许存在的偏倚究竟有多大?就电话调查而言,偏倚主要有三个方面:未登录电话的住户数、没有电话的家庭户数和只有移动电话的电话用户数。在美国的大多数城市地区,未登录的住户的百分比大大高于没有电话的住户。令人欣慰的是,我们在第8章将要讨论的随机数字拨号法(random-digit dialing,RDD),使研究者得以纠正用电话簿作为抽样框产生的、因为未登录的电话号码而导致的合格被调查人丢失的偏倚。然而,因丢失没有电话的家庭户而引起的偏倚并不会因此而纠正,要纠正这样的偏倚,我们必须花费一定的人力和物力,把这些人一一找出来。没有电话的住户的百分比的估计值,根据估计所依据的原始资料的不同会有所不同。根据2000年的普查资料估计,这一数字约占总家庭户的2.4%(U. S. Census Bureau, 2000a, Table DP-4),而2003年进行的当前人口调查(Current Population Survey,CPS)的这一估计数字为4.5%(Belinfante, 2004)[2]。此外,美国人口中,约有750万左右的人只使用无线电话,根本没有固定的住户电话(Carroll, 2003)。据估计,3% ~5%的移动电话用户已经完全不再使用固定电话,这些人是一伙年轻的、生活在城市中的、驾驶着汽车的专业人士。因此,非常有可能的是,全国住户中的6%根本就没有固定电话。所以,在一般情况下,为了找到一户没有电话或没有固定电话的住户,我们需要访问调查17户(16.6 ×6 ≈ 100%)——这需要耗费很多时间和金钱[3](记住,这是全美国的数字,就每一个特定的地区而言,这个数字可能不尽相同)。

大多数研究者在进行调查时,都忽视了这些因排除了没有电话的住户而产生的偏倚。他们之所以这样做,一般有三个原因:第一,他们认为没有电话的住户对最终结果的影响是微不足道或可以忽略不计的。因为有电话的群体在数量上大大高于无电话的群体,所以在来自无电话群体的结果能对最终结果有所影响之前,二者之间存在的差异必须相当大。第二,将无电话户包含进来的费用巨大,更何况,他们只占总人口的非常小的一部分。第三,总的样本结果才对研究者具有最重要的意义,而子群体的结果则不是那么重要的。然而在有些情况下,这些假设并不总是有效的,因而,对排除无电话户而可能产

生的影响,我们必须审慎考虑,不可贸然行事。

在美国各个地区,随地区和人口特征的变化,无电话户的比例也会有所不同。在 2002 年 3 月,美国西北部地区约有 3% 的住户为无电话户。而在美国的其他地区,在中西部地区、南部地区和西部地区,这一数字分别为 4.2%、5.5% 和 2.9% (U.S. Census Bureau, 2003)。收入水平与电话的覆盖率高度相关。在贫困线以下的住户,美国上述地区的无电话的数字分别为:9.6%、15.1%、16.0% 和 7.5%。甚至在有些州,以 1984 年的美元计,收入在 10 000 美元(相当于 2003 年的 17 954 美元)以下的住户中,电话的覆盖率低于 85%。其中:阿拉巴马 81.1%,阿肯色 82.7%,伊利诺伊 80.0%,肯塔基 84.7%,路易斯安那 82.8%,密西西比 82.4%,内华达 84.7%,新墨西哥 81.7%,奥克拉荷马 82.0%,弗吉尼亚 84.9%,怀俄明 84.4% (Belinfante, 2004)。拥有电话的住户的百分比随着教育、家庭收入和年龄的上升而上升。年龄在 65 岁及以上的住户中,电话的覆盖率最高。

我们必须考虑,在调查中是否的确需要把无电话户包括进来。同样,答案也是依我们研究的问题、可资利用的资源和研究进行的地点而异的。如果我们感兴趣的问题是失业者或那些生活在贫困指数以下的人们的观点,那么我们可能就应该考虑采用组合式的调查方法,或电话调查之外的其他调查方法。假如我们在 2000 年采用电话调查方法对这些人进行调查,那么每 6 户失业者家庭,就会有 1 户被排除在外,而每 4 户生活在贫困指数以下的家庭就会有 1 户被排除在外(我们应该记住,这一数字是全国性数字,而特定地区的这些数字可能与之显著不同)。我们同样也了解,无电话户都是低收入或正处在转变中的住户[4]。就我们所研究的变量而言,无电话住户的人们可能与有电话住户的人们对问题有不同的看法,行为方式也可能会有所不同。我们应当不时查阅我们调查的地区的人口普查或其他可资利用的数据,以使自己的研究计划得更为周密(参见第 7 章和第 8 章)。

设计问卷

我们需要从被调查人那里收集什么样的信息和如何最清楚地将我们的想法表达出来是调查设计阶段所必须尽早做出的两个决定。我们必须清楚我们问被调查人的问题是开放式的占多(如"什么是今天家庭遇到的最为困难的问题?"),还是封闭式的占多(如"您同意还是不同意进行流产?"),或是二者兼而有之。在前一种类型的问题中,被调查人用他们自己的语言来回答问题,而在后一类型的问题中,被调查人则需在事先所提供的答案的清单中进行选择。我们也必须确定,我们是否会问及有关态度、知识或行为方面的问题。与此同时,我们还必须确定我们需要什么类型的人口学信息,因为这一决定将会对我们采用什么方法来收集数据产生影响。如果问卷的问题形式基本上是①

一份答案清单，被调查人必须从中选择自己的答案；②给被调查人以其他可视的提示，以帮助他们给出答案（如一张我们希望被调查人进行评价的画片或一张刊有各类答案的卡片等）；或③要求被调查人查阅有关的个人记录，或做一些其他形式的、有助于他们回忆的工作，那么面对面的访谈就可能是我们的最佳选择。同样，如果我们需要问一些开放式的问题的话，那么邮寄式调查显然就不是一个好的选择，因为许多被调查人可能不会去完成一张需要写很多字的问卷。我们总是希望能使被调查人要做的工作变得尽可能地简单容易，从而最大程度地避免无回答情况的发生。

确定可以支配的经费和时间

在初步设计阶段，还有其他两个重要的问题必须考虑，那就是我们可用于调查的经费和时间的数量。经费和时间决定了我们可以做多少个调查；是否可以聘用调查员做调查和聘用编码员编码；如果可以，各可以聘用多少；在有些情况下，它们还决定了调查包括的地理范围能有多大。互联网调查是最快捷和最经济的，然后依次为邮寄式调查和电话调查，而面对面的调查是最为昂贵的。在有可能进行集体调查时，例如对一班学生进行调查，那么它的迅捷和经济程度可与互联网调查比肩。就时间而言，电话调查比面对面调查更为快捷，然而，完成这两种调查所需要的时间在一定程度上都取决于样本地理分布范围的大小。互联网调查和邮寄式调查者受这些变量的影响比较小，因为不论样本分布的地理范围大小如何，进行这两种调查的程序和方法都一样。就邮寄式调查而言，时间进度是固定的，一般不会受样本大小的影响。在第一份调查邮件寄出的两个星期之后，应该再寄一张目的在于提醒被调查人配合完成调查的明信片。再过两星期之后，寄第二封信和问卷；再过几个星期之后，可能需要寄第三封信或打电话查问。因此，不论样本大小和地理分布情况如何，邮寄式调查通常会持续 8 ~ 10 个星期。

分析数据和报告研究结果

在初步设计阶段，我们还必须考虑其他一些问题。首先我们应该考虑，为了回答研究提出的问题，我们需要选择哪些分析方法。然后则要考虑，准备研究报告、论文或其他由研究结果产生的成果所需要的时间。如果我们自己不能完成结束阶段的全部工作（如数据分析和解释、写作和打印），那么我们必须考虑由谁来分担这些工作，并给完成这些工作分配一定的时间。这些考虑之所以重要，因为它不仅是估计调查总费用所必需的，而且是计划并确定完成调查工作所需的人数和工时数所必需的。随调查工作的类型和范围的不同，调查的时间进度和费用的差别可能是非常大的。

步骤二:试调查

调查过程的第二个阶段是对调查的初步设计进行测试。在这一阶段我们需要做的工作有:准备好抽样框、保存记录的表格和调查的问题,然后对这些项目进行测试,以了解用它们来进行调查是否成功。首先,我们要收集整理抽样框,或采用更为便捷的方法,确认一个已有的抽样框是否适合我们的目标总体。

草拟问卷

在草拟问卷之初,从其他研究借用一些问题的方法不仅是可以接受的,甚至是应当鼓励的。在回顾那些过去已经做过的相同或类似的调查时,我们可以了解其他研究者是如何定义他们的概念和使用什么问题来测度这些概念。如果我们同意他们的做法,且这些问题和量表不存在版权问题[5],我们便可免费使用同样的措辞。如果我们只是部分地同意他们的做法,我们可以只使用我们同意的那些部分,然后再加上我们自己的观点和措词。使用他人的问题还有另外一个优点:采用与他人相同的问题,使我们得以将从我们调查区域内得到的调查结果与以前的研究结果做比较。

关于问卷设计,我们有两个忠告要给读者。第一个是,我们决不要因为我们使用的是他人使用过的问题,且使用者报告说他们在使用这些问题时没有任何问题,而想当然地认为将它们用于我们的调查的被调查人就必定也没有问题。不论我们从其他调查借用的问题有多少,我们都必须在正式开始收集数据之前进行试调查。试调查,好比正式演出前的彩排,是调查最为重要的组成部分之一。实际上,如果我们设计或提出许多新的问题,那么就可能需要进行两、三次试调查以保证被调查人能理解它们,从而使它们能成为一个微型的量表,收集到我们想要收集的信息。在我们对问卷感到完全满意之前,我们最好能不断地进行试调查——这可能需要做 30,40 或 50 次调查——如果在正式的数据收集工作开始之后再发现问题,就必须暂停问卷收集来纠正它们(这几乎是不可能的),或结束这一存在明显测量问题的调查。

第二个忠告与问卷的草稿数有关。问卷只有在经过许多稿之后,才有望解决它存在的所有问题,并使被调查人对问题的理解趋于一致。高质量的研究和粗制滥造的研究区别在于,前者在努力寻找或预见问题,而后者假设只要问题被回答便是有效的。我们应该坚决反对诸如此类的伪假设。

使用调查员

对于电话调查和个别访谈,我们还需要开始考虑使用多少调查人员的问

题。不仅如此，我们还必须开始起草和编写调查员培训材料，使调查员了解我们希望他们做什么。编写培训材料和培训调查员二者都是我们必须做的工作，无论我们使用的是专业调查员、志愿者，还是学生。

遴选和雇佣调查员的工作必须与样本的选择密切配合。对于个别访谈而言，我们希望调查员最好能生活在入选住户的附近，以最大程度地减少与被调查人联系所需的通勤时间和费用。例如，作为课堂作业项目，为了访问镇北的被调查人，我们应当分配生活在就近地区的学生到那里进行调查。而电话调查因为通常都在一个被集中监控的地点进行，所以调查员住在哪里并不重要。

分配给试调查的时间决定了试调查中需要调查员的人数。如果我们用一个星期的时间进行试调查，需要调查的人数是 30 个，那么我们可能需要聘用和培训 6 个调查员，每个调查员需要调查 5 个被调查人。而如果我们使用的是电话调查，那么我们可能需要聘用 3 个调查员，每个调查员平均需要完成 10 个调查。如果可能，我们应当使用最优秀的调查员，因为我们需要从他们那里得到有关我们问卷中可能发生的问题或调查方法和程序方面的批评性的反馈。

听取调查员汇报

在每次试调查结束的时候，研究者应该安排会见所有的调查员，听取他们的汇报。调查员的汇报应包括问卷中的所有问题和所有的调查步骤，发现存在的问题，并提出解决问题的建议。当某人的观点引发争论，或其他人可能遗忘什么想法时，小组汇报会应该鼓励相互进行交流。来自汇报会的反馈，加上审读实际调查记录将为问卷的进一步修改提供依据。

确定试调查的方法

试调查可以采用各种不同的方法和步骤进行。在最初阶段，试调查的目的在于收集对一个个问卷题目的反馈。而在以后阶段的目的则在于测试整个问卷和调查的方法步骤。问卷题目的测试，通常通过各种正式和非正式的途径。在我们开始撰写问题的时候，我们应该考虑到被调查人会对我们的问题做出何种反应，或做出什么样的解释。如果可以找到适当的总体成员，那么我们很快就能从他们那里得到他们对问题的反应。同事和学生也可用来作为测试问题的对象，听取他们对问题的意见。问卷的第一稿也可以非正式地在家庭成员、朋友或其他学生中进行。在经过非正式的试验之后，觉得问卷可以拿得出手的时候，我们便可以在调查地区的正式的调查对象中进行正式的试验。

在问卷的改进过程中，有两种方法现在已被越来越多地使用。一种是认知访谈，另一种是焦点组[6]。认知访谈是一种一对一式的交谈方式，即只有一个调查员和一个被调查人在场的访谈，当然访谈可能由若干对调查员和被调

查人同时但分别进行。调查员询问被调查人有关他们对所提出的问题的想法。此外，调查员常常追问被调查人对某一特定词汇涵义的理解，或要求他们对问题进行解释，并据此判断他们是否理解。例如，问卷也许包含着下面这样的问题：

- 您认为食品的安全性的含义是什么？
- 您个人认为食品的安全性有多么重要？
- 您认为生物技术的涵义是什么？
- 您怎样看待用遗传工程或生物技术改变食品问题？

作为试调查的一部分，我们想知道“生物技术”和“食品安全”这两个词对被调查人意味着什么。了解被调查人在回答问题时的认知过程，将有助于我们更好地撰写问题，从而收集到质量更好的数据。

另一种方法是使用焦点组。所谓焦点组是为了帮助我们编写问题或使调查更为有效，而集合起来讨论某些议题、问题或其他事情的一小群人。焦点组访谈有两种比较重要的用途。第一，在工具设计完成之前，焦点组访谈可保证我们在将要进行的工具设计中不遗漏重要的问题，如若不然，它也可使我们对将要开展的研究题目有更深更多的了解。例如，在低收入人群的住房问题研究中，研究者肯定已经对研究的题目有所了解，但可能很容易忽略其他那些尚未广为人知的或最近才出现的问题。第二，焦点组可作为一种测试问卷草稿的工具，特别是设计用于自填式问卷调查的问卷草稿的工具。在这个意义上，我们可以要求焦点组成员先根据自己的理解填写问卷，然后再对问卷进行小组讨论。一般讲，小组由 8 ~ 10 个具有相同特征的人组成：全部是男性、女性、年轻的黑人男性、工人等(Greenbaum, 1998)。小组的组成在一定程度上取决于调查的主题。小组的组成方式应该以能使小组成员在交谈的时候感到不拘束为宜。如果调查内容是有关性行为与感染艾滋病的关系，那么就在分组的时候，注意将男女分开就很重要。另一方面，如果调查的内容与种族问题有关，那么区分性别就远不如区分种族重要。

在近年的家庭和健康的研究中，有一个研究人员在设计问卷之前，进行了两个焦点组访谈。一个小组的成员都是单亲者，而另一个则包括有子女的已婚者。焦点组访谈的目的在于确定不同的家庭结构、家庭关系和生活经历对身心健康及家庭决策的影响，以及家庭内部活动和家庭成员关系对健康产生的影响。例如，小组成员被问到了这样一些问题：您如何应对您孩子身上发现的病状？谁带孩子去看医生？您是否向其他家庭成员请教？您的工作是否对您承担的家庭责任有影响？在您的家中，健康问题是否很重要？为了改善您的健康状况您做了一些什么？研究者特别感兴趣的问题是参与者之间存在的不同之处及他们用于描述他们的健康状况和保健行为的措辞和用语。

我们对问卷的草稿已经感到满意，或已经大致可以确定(尽管不能完全确

定）被调查人会如何回答某些问题，我们还是应该在那些真正的被调查人身上试验一下调查问卷。一个试调查通常要调查 20～40 个被调查人。而人数是否超过这一数目，则取决于以下几个方面：我们感兴趣的子群体数、这些子群体对问题的理解程度、我们对我们已经在多大程度上测量了想要测量的概念的把握程度、时间进度和经费预算等。在完成试调查并听取了调查员们的汇报之后，我们必须要对问卷和调查的方法步骤进行修改。如果改动的地方很多，那么我们必须先进行非正式测试，然后再进行另一次非正式的试调查。第二次试调查的规模取决于我们需要解决的问题不确定性的大小。在我们对问卷和调查的方法步骤感到完全满意之前，我们也许需要不断重复这一个程序。

试调查阶段可能延续一个月，甚至更长的时间。如果试调查既包括问卷的设计定稿，也包括培训材料的编纂，那么它就有可能延续两三个月。

步骤三：终结调查设计和计划

试调查的结果可以用于改进调查设计和执行计划。例如，在一个涉及四个医疗专业群体电话调查的试调查中，某一研究人员设计的问卷得到的回答率在 40%～50%。那么在随后与医务人员进行的焦点组访谈中，我们就可以集中讨论如何提高回答率的问题。在讨论之后提出的建议是：应该给医务人员在配合完成调查的问题上以选择的机会——他们既可选择填写邮寄式的问卷，也可接受电话调查。这样做的理由是，让每一位医务人员选择最适合自己工作日程的调查方法。这似乎是一个不错的建议，因为整个正式调查的回答率上升到了 67%。

试调查也可以在确定邮寄式调查跟踪联系的时间间隔，或与被调查人的最后一次联系是使用邮件还是电话等方面给我们以帮助。在抽样调查中几乎不存在什么确定而快捷的法则。为了提高回答率和数据质量，研究者必须因时因地制宜，随时做出各种调整。

在这一阶段，抽样计划、问卷、调查培训方法和培训材料、数据编码方案和数据分析方案等问题应做最后的调整。例如，在抽样过程中，我们可能会明白，为了达到我们希望的调查完成数，需要选取更多的电话号码，因为在试调查中，我们发现的空号数比预计的多。在调查问卷方面，我们可能会发现，如果改变一下问题的次序，或是把原来的开放性问题变为封闭性的，会使访谈过程变得更流畅。另外一个经常发生的问题是，我们可能会发现总体中的某一特定子群体，譬如年龄在 55 岁以上的人或黑人，他们的回答不同于其他子群体，或不同于我们的期望；或者我们可能会发现，某些子群体调查完成数低于我们所期望的。这时，我们需要确定这些子群的样本量是否足够，是否需要从

某些子群体超比例地抽取样本。

步骤四:数据收集

在这一阶段,我们需要对抽样的结果和数据收集过程进行监控,并开始编码和准备数据文件。在抽样阶段,我们希望确保所有的个案都是合乎规范地被抽取的,与此同时,我们也希望对所有这些个案的调查能顺利完成。电话调查和面对面调查中无响应的被调查人应该在本周的另一天或同一天的另一个时间再一次进行联系。对完成的个案的结果或处置情况的监控同样也是很重要的。拒绝接受调查、无法联系、不合格和完成调查的比率各是多少? 数据是否有作假的迹象? 一般来说,我们要对每一个调查员完成的工作的 10% ~ 20% ,通过督导再次与被调查人联系,和/或对完成的调查进行验证(cf. Johnson, Parker, & Clements, 2001)。在督导对被调查人进行回访的时候,不仅要了解调查是否确实进行过,而且还要了解进行的时间有多长,因为另外一种作假的手段是在调查过程中故意删掉某些问题。通常我们在回访时应向被调查人说明,这样做的目的是为了控制调查的质量。在电话调查中,我们对实际调查过程实行广泛地实时监听,以此来取代或补充回访。我们也可采用隔周或每周样本处理情况报告的办法,以便尽早发现和解决调查中发生的问题,避免问题成堆,无法收拾。

对调查员进行检测也同样重要。我们应该深入全面地检查每一个调查员最初完成的那几份问卷,以确证没有任何问题被遗漏,得到的信息是完整的,且所有关于调查的规定都得到了遵守。重要的是我们必须把检查的结果及时地反馈给每一个调查员,以便他们能在问题和错误成为定势之前予以纠正。我们应该对调查员对分配调查任务的完成情况进行检测,因为初来乍到的调查员很可能因轻信被调查人诸如"我实在太忙"或"我对调查的问题不感兴趣"这样拒绝接受调查的借口而放弃对既定个案的调查。如果我们能在培训阶段对这些问题进行讨论,并告诉他们如何应对,无疑对调查的成功是很有帮助的。如果一个调查员的拒访人数比例异常的高,那么我们就应当和那个调查员一起对那些个案一一加以检查,并找到对被调查人进行回访的方法或以后在遇到类似问题时的处理方法。优秀的研究人员总是想方设法通过使用训练有素的调查员对拒访者进行回访,并促使他们转变态度接受调查,从而最大限度地减少持不合作态度的人数。通常情况下,初次拒访者能转化 20% ~ 40% ,格罗夫斯和库珀(Groves & Couper,1998)提供了一些减少拒访或无法联系的方法和技巧。

在完成数据收集之后,我们必须对它们进行编码,并把完成的调查信息做

成电脑数据文件。这些工作统称为数据简编,通常它由另外一组人完成。在将数据送入数据文件之前,数据简编人员必须首先对问卷进行编辑整理。他们将在问卷中寻找调查督导曾经寻找过的那些同样的东西:略过的问题、不完整的信息以及不准确的、前后不一的或不符合逻辑的条款。这种编辑整理,应在调查完成之后尽早进行。比较理想的是在调查完成后的几天之内进行,一般不要在两个星期之后才去做这项工作。时间问题是很重要的,因为为了解决问卷中发现的问题,我们需要与调查员、被调查人联系。联系得越早,当事人越容易回忆起当时的情况,从而也就越容易使问题得到解决。

步骤五:数据编码、数据文件的建立、数据分析和最终研究报告

调查的最后一个步骤包括编码、数据分析和撰写最终的研究报告或论文。所谓编码,就是将被调查人对调查的问题给出的答案标上数码[7]。对被调查人给出的每一个问题的答案进行编码,从而使我们得以进行特征值的估计,或在变量中寻找它们的关系模式。下面的例子形象地展示了编码究竟是如何进行的。

在一个马里兰州成人对街头暴力犯罪问题的态度的电话调查问卷中,共回收问卷 824 份(见附录 B)。我们对被调查人给出的答案,按一个一个被调查人和一个一个答案,一个一个地进行编码,然后再按人将编好码的答案,按一个一个被调查人,一条一条地送入数据记录(每一个被调查人给出的所有答案为一条记录,译者注)。一般讲,答案在数据记录中排列的顺序与问题提出的顺序是相同的。计算机辅助电话调查和纸笔式调查的编码和录入方法没有什么差异。一条数据记录包括了一个被调查人所有的编好了码的答案。此外,每个被调查人都有一个 001 ~ 824 的唯一的身份识别数码,通常它都被输入在每一个被调查人的数据记录的最开始的几个位置。

在编码时,每一个问题或变量在个人数据记录中占据的空间位置是事先规定的,即占据的列数和这些列在记录中的位置(第几列)都是相同的。每一个问题或变量、每一个答案类别都给出了一个规定的数码。在马里兰州的犯罪调查中,第一个问题是有关发生在被调查人邻里地区的犯罪问题。这一问题的答案共有四类,所以编码为 1 ~ 4,再加上一个“不知道”的编码 8。如果 ID 为 001 的被调查人给出的答案是“不太严重”,那么它的编码就是 3。如果对第二和第三个问题,给出的答案为“很严重”和“几乎没有什么变化”,那么它们的编码分别为 1 和 3。如果数据记录以 ID 数码开始,那么第一个被调查人数据记录的前六列的编码值就是 001313。我们可以用与此相同的方式,将数码分配给第一个被调查人的其余 49 个问题的答案和其余 823 个被调查人

给出的问题答案。在编写和改进问题答案的分类时,有一条简单的法则我们应该牢记:答案分类总是应当互斥和穷尽的。

在开始数据分析之前,我们应当对数据进行检查或"清理",以识别并改正编码和数据录入过程中发生的错误[8]。在数据清理过程中,对每一个问题的答案或变量的答案的编码都必须检查是否存在非法代码,并在可能的时候,对相关变量进行跨变量的一致性检验。如果我们规定被调查人性别的编码为 1 男,2 女,那么我们在数据清理过程中,便要将这一变量中所有码值为 3 或大于 3 的记录找出来,并一一加以改正。跨变量的一致性检验的方式也与之相似。如果某一组问题是适用于男性的被调查人,那么女性的被调查人的所有这些问题的答案都应当是"不适用"。我们一般需要在最后一个调查完成之后的几天到 4 个星期的时间完成包括编码、清理和准备数据文件在内的所有的数据简编工作。时间的长短取决于数据收集的方法和调查的复杂程度。在使用计算机收集数据的时候,这些工作已经"前移",成为了问卷设计和改进的一部分。

数据分析和研究报告撰写所需时间的长短,往往要视调查目的和承担任务的性质而定。一般讲,在一份无效代码被清除过的数据文件准备好了之后,我们需要 4 ~6 个星期来撰写调查报告。报告的内容包括调查的方法和过程及对一个或几个我们所研究的问题给出的回答。如果我们的计划更宏大,那么撰写报告需要的时间可能更多一些。

研究时间进度表实例

前面有关调查基本步骤的介绍,使我们对以下问题已经有所了解。第一,我们已经了解进行一个调查必须要就某些问题进行决策,而在进行这些决策时,有多个方面的因素是我们必须要考虑的。第二,我们还了解了这些性质不同的工作需要同时安排进行。第三,我们还对调查中某些工作所需要的时间长短有了大致的了解。在下面的例子中,我们将更为具体地给读者介绍一个特定的时间进度表。该进度表是为一个全美 RDD 调查设计的。该调查共调查了家中有 18 岁及以上成员的家庭 1 000 户。调查的主题与生物技术和食品安全有关(表 2.1 列出了该调查的主要任务和每项任务的时间分配)。在下面的讨论中,我们所讨论的调查涉及的关键工作,都是基于计算机辅助的电话调查(Computer-assisted telephone interviewing,CATI)的。

首先确定调查涉及的范围是整个美国。进而研究者又确定,对于这一调查比较可行的方法是 RDD 的电话调查。在这两个问题确定之后,我们必须解决的问题是拟定研究的问题的草稿,确定需要被调查人回答的问题的具体类型。而在进行这些工作的时候,我们需要检索那些期刊文章和书籍中与我们的研究题目有关的文献。在阅读相关文献的时候,我们应该把注意力放在研

表 2.1　1 000 户随机数码拨号电话调查的时间进度表

工作内容	需要的时间/周	起止时间/周
文献检索和起草问卷	8	1 ~ 8
征求 10 名被调查人对问卷草稿的意见	1 ~ 2	8 ~ 9
修改问卷	2	10 ~ 11
进行第一次试调查(n = 25 ~ 40)	1	12
听取调查员的汇报,修改问卷	3	13 ~ 15
进行第二次试调查(n = 20 ~ 30)	1	16
听取调查员的汇报,修改问卷和编写培训材料	4	17 ~ 20
选取样本(为试调查和正式调查)	12	8 ~ 19
进行正式的数据收集	8	21 ~ 28
数据编码和准备数据文件	12	21 ~ 32
分析数据和撰写报告	开放,时间不限	开放

究者是怎样定义和测量他们的概念的。与此同时,还必须关注他们在问卷中使用的具体题项,尤其要关注的是什么样的做法成功了,什么样的做法没有成功。由此得到的各种信息都可用于我们自己的问卷设计。虽然我们给开始阶段的文献检索和以往研究的回顾分配了 3 个星期的时间,但是为了解决遇到的各种问题,对新文献或新观点的关注应该贯穿整个调查研究工作。接下来的是把问题组织成一张问卷。我们可能需要 4 ~ 5 个星期的时间来编写问题,修改问卷。我们应该争取与几个同事或朋友讨论每一稿,做一些修正,并在不同的人中进行测试,做一些修正。在修正之后应该再次进行测试。在这项工作进行的同时,那些负责样本选择的人则必须开始收集、编制,或从学术或商业机构那里购买抽样框。

因为在生物技术领域内做过的社会科学研究不是很多,我们对公众在这一方面的知识和态度的了解不可能很多。因此召集几次焦点组访谈,或进行几次边想边说的访谈,无疑将有助于我们的问卷设计。我们给这一工作分配了 1 ~ 2 周的时间。在焦点组访谈或边想边说访谈进行之前的 2 个星期左右,参加者就应当确定下来了。

在第 10 ~ 11 周,我们将再一次对问卷进行修改,为第 12 周的正式试调查做好准备。试调查基于一个全国性的一般人口样本进行。因此,在第 12 周之前,我们必须完成调查员的招聘,并编写和印制好培训材料。在试调查完成之后,我们应该立即安排听取调查员的汇报,这一工作也许需要半天时间。如果可能的话,我们应该通过汇报会找出问题,并讨论如何解决它们。如果访谈是

用 CATI 进行的,那么日程安排中要考虑的关键问题是 CATI 软件的性能和那个我们指定的将问卷装入计算机系统的人的经验。某些 CATI 系统,譬如 CASES(加利福尼亚大学,伯克莱分校设计)使用的编程语言,要求使用者有一定的编程经验。其他一些系统(如 Sawtooth)虽然不能处理比较复杂的问卷,但却比较容易使用。重要的问题是系统和程序员之间必须高度相配。

与初始编程密切相关的问题是调试和重新编程。调查的工具一旦装入系统,需要马上进行"筛试"。所谓筛试,就是我们需要有一个专门的人在安装好的系统上对调查所需要经过的每一路径试行,以确保系统的运行能符合我们的要求。一般讲,我们总是会筛试出一些问题,因此我们需要安排一定的时间来纠正发现的问题。

因为已经料到将会对问卷做比较重大的修改,所以我们计划进行第 2 次正式的试调查,被调查的人数在 20 ~ 30 个。问卷的修改和第 2 次试调查的准备工作被安排在了第 12 ~ 15 周,第 2 次试调查被安排在了第 16 周。另一次调查员汇报会则被安排在了这一次试调查结束之后。

这一次试调查和汇报会可能没有发现什么重大或令人吃惊的问题,或者可能发现问卷需要做一些修改,还需要再安排进行一次试调查。根据问题的复杂性或设计本身的需要,可能需要再安排一次试调查,在遇到这样的情况时,我们大可不必因此而大惊小怪。假如我们假定一切都进展得很顺利,那么在接下来的 3 ~ 4 周(第 17 ~ 20 周),我们将完成问卷的最后一稿和培训材料。与此同时,我们需要着手准备选聘和培训编码员,编写编码的参考材料和数据清理及分析文件的说明。在第 21 周之前,问卷和相关材料必须印制完成。因为我们计划完成 1 000 个个案的调查,所以我们可能需要印制 1 400 ~ 1 500份问卷。如果每本问卷 20 页,1 450 份问卷本身就需要 29 000 张纸(58 令)。我们要分配足够的时间给定稿问卷的打印、复制印刷、分页和装订等各个工序。然后,我们还须把培训材料、问卷和样本分配表及其他调查管理材料整理在一起。

如果我们采用 CATI 调查,第 17 ~ 20 周之间安排的那些工作仍然是必要的。此外,在试调查和问卷修改之后,修改过的工具仍然要用 CATI 进行测试。这是一种与问卷内容测试完全不同的另一种方式的测试。这不仅需要安排专门的时间,也需要专门的技巧。如果试验缺斤少两,很不充分,那么数据收集工具可能就难以使用,调查就有中断的危险,而我们因此所付出的代价将是十分巨大的。如果我们原来没用使用过电话调查系统,或缺乏有经验的程序员,那么我们就应该在调查设计和执行的时间外,专门安排时间来学习调查系统的安装、调试和操作。在使用计算机辅助的个人访谈系统(computer-assisted personal interviewing,CAPI)和其他的计算机辅助访谈系统(computer-assisted interviewing,CAI)时,我们同样也要注意这一问题。这些系统功能固然

很强大,但为了有比较高的使用效率,我们需要有这方面的专家。

在第21周,我们将开始培训调查员,并进行数据收集。我们安排了8周时间来收集数据。我们安排的时间必须足够,这样我们才能有机会对每一样本个案进行调查。我们需要有足够的时间来联系那些外出度假、忙得不可开交、经常不在家、生病或临时住院治疗的人们。每一个个案都必须能完全按规定进行处理,这样样本才能充分代表总体的各种性质。此外,我们必须也要考虑调查员的问题。我们可以要求调查员每周工作多少小时?有多少调查员平均每周工作的时间能达到我们的要求?如果他们都在中央电话中心进行调查,我们是否有足够数量的个案满足他们整个工作时间的需要?如果我们聘用了15个调查员,要求他们每周平均工作15个小时,且假定在整个数据收集阶段,没有人中途退出,那么每个调查员每周必须完成8.3个调查。如果没有什么重大的问题发生,那么这个目标应该不难达到,整个数据收集工作将会在8周之内全部完成。

在调查员培训的那一周,我们应该着手进行编码员的培训和分析文件的有关规定的编写。编码应和调查同时进行,因此,要求编码员应尽早对完成的问卷进行编辑整理。我们的进度表在数据收集工作完成之后,安排了4周时间来完成编码、数据清理和准备好数据分析文件的工作。如果问卷中只有一或两个开放式问题,且又由调查员完成全部的后期工作,那么从编码到最终准备好数据分析的文件,可能不需要4周。

完成一个调查所需要的时间长短,取决于几个方面的因素,其中包括项目的复杂程度、样本的大小、收集数据使用的方法、收集数据的数量和复杂程度、对数据分析的具体要求及其他可能需要考虑的因素等。例如,回答率可能会受收集数据实施时间的长短的影响。回拨的次数越多,原来拒绝接受调查的被调查人态度转变的可能就越大。诸如这样的策略,也同样适用于邮寄式调查,邮寄促使拒访者转变态度的信件,特别是包含某种激励手段的信件可能颇为奏效。但所有这些手段都需要一定的时间。在其他条件相同的前提下,调查实施的时间越长,采取各种措施提高回答率的可能就越大。做一个高质量的调查,自始至终,不仅需要经过很多阶段,而且还需要在各个阶段根据具体情况做出不下数百个的决定,因此是很费时的。在我们列举的例子中,研究思想的概念化、调查项目的实施和数据文件的建立,前后共花费了8个月的时间。这大致反映了规范地、不删减任何程序地执行一个中等规模的调查所必需付出的时间和努力(这一时间表不包括数据分析和报告撰写)。然而,我们并非总是能有那么充裕的时间来实施一个调查。在有些情况下,我们需要在比较短的时间内完成一个调查项目:需要确定公众对某一新闻事件的反应(如军事入侵或企图暗杀总统),或要求一个调查作为一个可当作教学任务来完成。

在时间比较仓促的时候，研究者必须：①认识到简化调查步骤可能有某种影响结果的信度和效度的缺陷；②确定我们应该采取什么折中的办法，以使这样的影响降到最小。例如，在必须采取折中措施时，研究者可能希望到某个商业机构去购买样本，而非自己亲自设计和抽取样本；或研究者可能希望使用以前问卷中用过的题目，以在自己的调查中节省为确定题目的信度和效度而进行的试调查的时间。在这两种情况下，研究者都只是假定事情将会顺利进展，因而不管愿不愿意，他们都必须要冒一定的风险。

注 释

1 我们在第 8 章将会了解，有一种更好、更为简单的抽样框可用于这一类型的电话调查。

2 导致这些差异的一个原因可能是问题的措辞：人口普查的问题是“在这所房子、住宅或活动房屋内是否有您可以拨打或接听的电话设备？”CPS 问的是一组问题，共有 6 个问题。其目的是使问题 2 ~ 4 和 8 ~ 6 能通过电话进行调查。它的基本问题是“因为包含在这一调查中的住户将要接受调查（在随后的三个月中，还将一而再，再而三地接受调查），我们希望通过电话进行一个追踪调查。在这一住房/住宅内有电话吗？”

3 一般讲，手机的交流方式与座机是不同的。大多数 RDD 调查都不包括手机。其原因在于：①费用问题——在使用手机时，打电话、接电话都要付费，因此大多数手机用户都不愿意自己付费接受调查；②责任问题——存在一个人们担心的问题，那就是被调查人在接受调查的时候正在开车，于是有可能因此而发生交通事故，那样调查的组织者可能要担负一定的责任；③抽取的概率问题——手机的样本由个人组成，而座机的样本则由多个个体的住户组成。在把两种数据合并使用的时候，必须进行加权处理。我们将在后面的有关章节中讨论这一问题。

4 住户处在一种变动之中，这些变动包括成员的临时迁居、移民、失业、离婚、分居或丧偶等。

5 大多数研究者并不鼓励或支持调查问题有版权的做法。约定俗成的规矩是，为了测验题目的信度和效度，鼓励对不同的总体在不同的时间，任意使用那些相同的问题。

6 在第 6 章，我们讨论测试问卷和对结果进行评估的其他方法，如专家组、事后调查、行为编码的访谈录音和被调查人报告。

7 有的时候，字母或符号被用作答案的编码。这种情况是一种例外，而非常规，这样的编码不能用来做统计分析。我们对这一类的编码略而不谈。

8 在将数据直接送入计算机的时候，使用计算机辅助的电话调查或个别访谈，在问卷定稿之后和正式调查开始之前，有关编码规则和数据清理的具体要求都已经编写完成。这样，数据清理和编码的检查便成为了调查的一部分。

选择数据收集的方法

第3章

Selecting the Method of Data Collection

在确定抽样调查的确适合用于我们所要研究的问题之后,接下来我们所要考虑的问题便是究竟什么样的数据收集方法最为合适。迄今为止,三种最为普遍的收集数据的方法是邮寄式调查、电话调查和面对面访谈。而在过去的十年间,由于计算机辅助调查的崛起,使通过互联网(World Wide Web)进行的调查,大有后来居上,发展成为第四种调查方法之势[1]。我们也可以把这些调查方法组合起来使用,这一问题将在本章的稍后部分进行讨论。第五种方法是问卷集中填写调查法(group-administered questionnaire),例如一张由入选的教室或班级中的所有学生在一个特定的日子共同填写的问卷便属于这样一种性质的问卷。这样一种类型的调查通常都需要辅之以其他的调查方法,我们也将在稍后讨论组合调查方法时对它进行讨论。下面我们首先对四种主要的调查方法做一个简要的介绍。

四种主要调查方法优缺点的评价

不存在什么"最好"的调查方法,每一种方法都有自己的长处和不足。重要的是,我们不但应该对它们了然于胸,而且能结合自己的研究目的,对每一种方法的优点和缺点详加考察。有关数据收集方法的讨论必须通过一个一个研究案例进行。在选择数据收集的方法时,我们在考虑三类主要因素的同时,还必须考虑每类因素涉及的许多次要的因素:

1. 管理或资源因素。
2. 问卷因素。
3. 数据质量因素。

表3.1对邮寄式调查、互联网调查、电话调查和面对面调查的优缺点进行了概括和比较。在确定调查方法时,我们所要考虑的问题涉及有关因素中的每一类。就管理和资源因素而言,我们必须考虑做一项研究需要多少时间、聘

表 3.1 主要调查方法比较

调查涉及的问题	邮寄式问卷	互联网调查	电话调查	面对面(在家中进行的)访谈
管理、资源因素				
费用	低*	**非常低**	低/中等	高
收集数据时间的长短	长(10 周)	**非常短/短(1～3 周)**	**短(2～4 周)**	中等/较长(4～12 周)
样本的地理分布	**可以比较广**	**可以比较广**	**可以比较广**	必须比较集中
与问卷有关的问题				
问卷长度	短/中等(4～12 页)	短(短于 15 分钟)	中等/较长(15～35 分钟)	**较长(30～60 分钟)**
问卷复杂程度	必须很简单	**可以比较复杂**	**可以比较复杂**	**可以比较复杂**
问题复杂程度	简单/中等复杂	简单/中等复杂	必须简短而简单	**可以比较复杂**
题序控制	差	差/一般	**非常好**	**非常好**
开放式问题的使用	差	一般/差	一般	**好**
可视辅助手段使用	好	**非常好**	一般不可能	**非常好**
住户/个人记录使用	**非常好**	**非常好**	一般	好
亲近度	一般	差/一般	好	**非常好**
敏感问题	**好**	差/一般	一般/好	一般(好,使用 CASI)
无胁迫性问题	好	好	好	好
数据质量控制问题				
抽样框偏倚	一般较低	低/高	低(采用 RDD)	低
回答率	差/好	差/好	一般/好	**好/非常好**
回答偏倚	中等/较高(比较适于文化程度较高的被调查人)	中等/较高(比较适于文化程度较高的被调查人)	**低**	**低**
对拒访和无法联系的了解	一般	一般	差	一般
对调查现场的控制	差	差	一般	**好**
记录的答案的质量	一般/好	一般/好	**非常好**	**非常好**

* 黑体字表示该方法在调查的各个组成部分的这一个方面,优于某个或其他所有的方法。

用调查员和/或编码员、购买硬件、软件和补给物品需要多少钱、是否需要使用激励机制，以及是否需要购买或构建一份我们准备抽取样本和进行调查的总体的清单等。问卷因素包括为了精确的测量研究概念和达到研究预期的目标，需要问多少问题，这些问题是什么样的。我们必须确定某种数据收集方法是否比其他的方法费用更高，或产生的报告误差更小。而在考虑数据质量时，我们必须了解，较之其他的方法，某种数据收集方法是否更容易取得被调查人的合作，如果我们使用它，是否能从被调查人那里得到更为精确或更为完整的数据，以及是否能更全面地包括我们希望对之进行研究的总体等。

尽管许多研究开始都想从邮寄式调查、互联网调查和电话调查中选择可行方法。但是，始终存在着许多不以我们的意志为转移的客观理由，如需要向被调查人显示那些关键性问题的答案清单、目标总体的电话或计算机的拥有率太低，或被调查人存在阅读和书写障碍等，使我们不得不采用面对面的访谈。我们必须首先回答的两个问题是："本次调查的被调查人是什么人?"和"我所研究的问题是否更适合采用某种方法?"如果对第二个问题的答案是否定的，那么时间和费用则是随后应该考虑的问题。在时间和费用之外，研究者还必须考虑若干个其他的重要因素（如表 3.1）。在下面几节，我们将参照表 3.1 所列的各种因素，讨论每一种方法的优缺点。

邮寄式调查

通常，邮寄式调查都需要预先给抽出的人或地址寄出一封简短的通知信，然后再寄去带有详细说明的封面信件和调查问卷（Dillman, 2000）。详细说明的内容应该包括：调查目的、调查的资助者和组织者、问卷应由什么人填写、调查的重要性、关于对调查材料保密性的保证和问卷寄回的具体时间（一般在 10 天之内）等。邮寄式调查的问卷必须明白易懂，语言必须清楚、简洁，不可夸大其词，因为各种不同的被调查人对说明和问题的理解必须一致。如果被调查人不明白要做什么，或觉得很困难，太费时间，他们就可能会无法或不愿意填写问卷，或虽然填写了，却填得不对。封面信应该提供一个电话号码，以便被调查人在对调查的合法性存有疑虑或对问题不太理解时和我们联系。然而，经验告诉我们，尽管我们给被调查人提供了联系电话，即使调查的内容颇有争议或涉及个人的某些信息，但真正给我们打电话的人却不是很多，可能不到 1%。当被调查人对调查心存疑虑或有顾虑时，他们更可能采取的做法是不填写问卷或跳过那些不清楚的题目，而不是打电话询问。那些为提高回答率而精心设计的调查，都会给没有寄回问卷的被调查人寄去明信片表示感谢，然后再寄去一份封面信和一份问卷。如果还是没有回音，最后可能还会通过

电话、快邮(Priority Mail)或特快专递(FedEx)进行"特别的联系"(Dillman, 2000)。我们还发现,在最初的信件中采用物质的和非物质的奖励有一定效果(Church, 1993)。

邮寄式调查的优点

邮寄式调查的费用,远远低于电话调查或面对面调查。它所需要的费用只包括邮资、购买信封的费用;收集或编制研究总体姓名和地址清单(抽样框)的费用;打印和印刷专业水准的问卷的费用;雇用整理、封装和邮寄材料,追踪问卷回收情况,对无回音的被调查人再次寄发邮件,对回收的问卷进行编辑整理和编码,将数据输入计算机文件和构建分析文件的各种工作人员的费用等。

邮寄式调查和电话调查或面对面调查的主要区别在于:邮寄式调查的数据不是由调查员收集的。这一特点既是一种优点,也是一种缺点。它最为明显的优点体现在调查费用上。邮寄一份附有回寄邮资和信封的 8 页问卷(印刷在 4 张纸上)邮件,即使使用头等邮件,邮资也是非常有限的,而一个调查员每小时的酬劳至少要 8 美元。邮寄式调查的另一个优点源自邮资的等价性。即在邮寄问卷时,被调查人无论远在阿拉斯加还是近在本镇,邮资都是一样的。因此我们可以用较低的费用进行一个全国性的抽样调查。不言而喻,全国性的样本肯定比城市范围或州范围的样本更具有多样性。

邮寄式调查使用的问卷也有几个优点。一个优点是被调查人可以查阅家庭或个人记录。例如,在我们需要了解被调查人在过去六个月中支付了多少美元的房贷利息或支付了多少医药费时,我们可以要求被调查人根据有关记录而不是记忆来回答问题。这样得到的答案必然更加精确。第二个优点是我们可以使用可视性辅助手段。如果我们希望被调查人在回答问题之前先看一份可供选择的答案清单,那么我们便可以将答案与问题一起列出。如果我们希望被调查人确定他们居住的邻里地区的边界或上班路线,那么我们就可以在问卷或附页上绘制一张地图。

邮寄式调查在收集敏感问题的数据时也比较成功。被调查人在回答那些羞于示人的生理卫生方面的问题(Hochstim, 1967)、个人破产、因酒后驾车获罪(Locander, Sudman, & Bradburn, 1976)和其他令他们感到窘迫的问题时(Aday, 1996; Tourangeau, Rips, & Rasinski, 2000),自填式问卷会使他们觉得放松一些。对于诸如这样一些类型的问题,数据收集方法的匿名程度越高,回答问题的比例就越高。然而,我们必须记住,尽管邮寄式调查在这些方面的确优于其他调查方法,但是它还远不是尽善尽美的。对许多敏感行为,即使我们采用匿名的邮寄式调查,它也同样无能为力(Sudman & Bradburn, 1982)。

不论样本大小和地理分布情况如何,进行一项邮寄式调查所需要的时间

一般是比较固定的——通常是8~10周。诚如前述,我们必须留出足够的时间,使问卷能寄达被调查人,保证他们能有充分的时间填写问卷并将问卷寄回给我们。此外我们也需要有时间对没有回音的被调查人再次发邮件和回收邮件。邮寄式调查需要的时间相对于其他调查方法而言,究竟是否有优势,要视具体的调查而定。如果我们的样本为2 000或多于2 000,那么邮寄式调查所需的时间可能少于电话调查。但在一个只有300~500个被调查人的调查中,电话调查一般只需要2~4周,而邮寄式调查则可能需要8~10周。

在我们讨论邮寄式调查的回答率为什么会变化很大这一问题之前,我们必须对回答率这一术语加以定义。我们将回答率定义为完成问卷填写的合格样本成员数除以合格样本成员的总数[2]。合格一词对定义至关重要,因为有时那些我们不希望包含在我们的调查中的人,可能会收到并填写一份问卷。例如,假定我们希望测量大专院校的男学生对流产问题的态度,如果我们将一份问卷寄给了一位女学生,且她也做了回答,那么我们便应将这份答案视为不合格。在计算回答率的时候,我们就不应该把它计算在内,应该在样本数中减去1。回答率是调查质量的重要量度,一般讲,回答率越高,质量越好。不过,30%的回答率比20%的回答率也强不了多少,二者都太低,都无法接受。

在调查的主题正是被调查人很关心的问题时,邮寄式调查的回答率可能会达到比较高的水平。例如,在一个有关玩忽职守的保险费费用负担和政府对此可采取的行动的调查中,如果调查对象都是医生,那么回答的情况就可能非常好。而同样的调查如果在一般人口总体中进行,结果就可能不怎么理想。其原因在于,在一般人口总体中,尽管的确可能有许多人认为这一问题非常重要,但他们对这一问题的关心程度远不如医生。

对提高邮寄式调查的回答率的确有成效的两种手段是采用一定的奖励和坚持不懈地与被调查人反复联系(Yammarino, Skinner, & Childers, 1991; James & Bolstein, 1990)。可供我们使用的奖励办法有很多种,包括数额不等的现金或支票、书籍、铅笔、钥匙链、领带夹、彩票和抽奖卷、慈善捐款和其他物质或口头承诺等。根据丘奇(Church, 1993)的有关报告,预付的货币或非货币的奖励的确可以提高回答率,而二者相比,货币奖励的效果更好,二者最多可分别提高回答率19.1%和7.9%。采用奖励的办法来提高第一次问卷的回寄率是有条件的。如果使用不当,可能不会有任何作用。关键在于必须在使用奖励办法的同时提出我们最基本的要求[3]。

邮寄式调查的缺点

在某一子群体较其他各个子群体更好或更不配合调查时,就会发生所谓的回答偏倚问题。例如,邮寄式调查就比较容易产生回答偏倚,因为在邮寄式调查中,有两类人群的回答率比较低:一类是文化程度比较低、阅读和回答问

题存在一定困难的人;另一类是对调查的题目不感兴趣的人。邮寄式调查和互联网调查的回答偏倚比其他的调查方法要大,因为在这两种调查中,被调查人更容易置问卷于不顾,而在一个彬彬有礼但坚持不懈的调查员面前置问卷于不顾却比较困难。正因为如此,封面信件和问卷的外观对鼓励合作和最大限度地缩小回答偏倚是至关重要的。一般讲,我们只有通过这些材料才能和被调查人取得联系,所以它们不仅必须非常有说服力,而且必须很有吸引力。

为了对可能产生的回答偏倚做出评估,我们必须尽可能多地掌握有关无回答(拒绝回答和无法联系)的情况。重要的问题是从因变量的角度确定那些拒绝接受调查和无法联系到的被调查人与那些配合调查的被调查人是否存在差别[4]。当我们掌握了总体清单的时候,我们就可以据之确定可能的回答偏倚。例如,通过对清单上的姓名和地址的分析,我们就有可能确定:男性与女性究竟哪种性别比较合作?生活在大城市中的人比生活在其他地区的人究竟更难合作还是更容易合作?我们手中的清单可能还含有其他一些有助于我们确定回答偏倚的信息。

邮寄式问卷本身固有的一些特点使得它比由调查员填写的调查效率要低。因为被调查人在决定是否填写问卷之前可以看到整个问卷,所以问卷不能太长,也不能太复杂或难以填写。与电话调查相比,在邮寄式调查中,无回答与被调查人对调查题目是否有兴趣的相关性更强。因为在电话调查中,被调查人在对调查内容有更多的了解之前已经决定是否参与调查了。

邮寄的问卷必须在自行了解的基础上填写,因为在被调查人对问题感到迷惑不解时,没有人可以在他面前提供帮助。即使是很有经验的研究人员,在他们自以为问卷表述的内容和提出的问题是不需要解释就可以理解时,也常常会有相当一部分的被调查人无法正确理解。而被调查人对问卷缺乏理解势必会影响调查结论的质量。被调查人很容易跳过那些他们不理解或不想回答的问题。此外,研究人员实际上很难控制被调查人回答问题的次序,甚至问卷的实际填写人。最后,与由调查员操作的调查相比,邮寄式调查得到的开放式问题的答案往往都不够完整和详细。除非有调查员的督促,否则许多被调查人给出的答案所提供的信息一般都是问题所需要的最低限度。这种倾向在被调查人文化程度较低或存在书写困难时尤其严重。有关在开放式问题上,自填式问卷和调查员填写的问卷之间在被调查人给出的答案方面存在差异的经验研究并不多见。有两种可检验的假设可能有一定的实践性。对于自填式问卷而言,人们也许期望问卷更长,更复杂,因为在填写那些问卷时,并没有时间的限制。而对调查员填写的问卷,我们也可做同样的假设,因为这有助于探究被调查人的基本观点(Groves et al., 2004)。德里(de Leeuw,1992)所做的一项研究发现,在自填式和调查员填写的问卷之间,四个开放式问题的答案没有什么差异。将来有关这方面的研究需要解决的问题是,在控制社会经济地位

的前提下，不同的数据收集的方法得到的数据，其质量是否会受被调查人对调查题目的兴趣的影响。

互联网调查

互联网调查是一个比较新，但却被越来越普遍采用的自填式问卷调查方法，它与邮寄式调查固然有许多相似之处，但也存在许多比较大的差别。由于覆盖面较小——到2003年为止，能上网的美国人只占总人口的一半（Federal Communication Commission，FCC，2003）——借助互联网进行的概率抽样调查的抽样，一般都以已知或可能的互联网用户的清单为依据。依据抽样框清单所能提供的信息详略不同，可能的被调查人可以通过电话、普通邮件或电子邮件进行联系，以告知调查组织者的身份、承诺对调查资料的保密，并对如何上网接受调查做必要的指示。为了保证使只有那些被抽到了的个体才能上网接受调查，且只接受一次调查，每一个被调查人都给定了一个唯一的个人识别码（PIN）。在上网接受问卷调查前，他们必须首先输入这一识别码。识别码既可附在最初接触的信件中，也可以作为扩展名植入在一个电子邮件指定的信息位置（URL）的扩展部分。在后面一种情况中，被调查人只要在网上点击信息位置，互联网的网站就会自动识别识别码，无需被调查人从键盘输入身份识别码，从而避免了可能的输入误差。

屏幕上首先出现的是一个相似而简短的导言或载有互联网问卷欢迎词的页面。通常，该页面会对调查的目的做一个简单扼要的介绍，对参与者进行必要鼓励。如果可能，该页面还可附如何输入身份识别码的简明指示。在这一屏之后，屏幕才显示问卷的第一页。欢迎词页面也应该包括可以提供帮助的电子邮件地址和/或电话号码，以便那些希望就调查的问题提问或在回答问题时有困难的被调查人与我们联系。将我们的邮政地址列在页面上也不失为一种明智之举，因为它可以为那些喜欢先将问卷打印出来，然后填好邮寄给我们的被调查人提供方便。在线填答的问卷的设计必须慎之又慎，它必须能为那些有不同文化程度和不同使用计算机能力，且使用不同的计算机硬件和软件及不同的上网途径的被调查人所能进入和理解。想方设法采取各种联系手段，提醒那些还没有登录调查网站或没有在规定的时间内完成问卷的被抽到的个人登录或递交完成的问卷是提高回答率最重要的方法。

互联网调查的优点

互联网调查的两个最大的优点是费用低和收集数据的速度快。互联网调查省去的不仅是面对面调查和电话调查的访谈费，而且还省去了邮寄式调查

的纸张、问卷复制、邮资和数据录入的费用。不仅如此,样本的地理分布和大小对互联网调查也没有影响。通过互联网收集一个全国性、甚至国际性的数据,其费用并不比一个收集地理位置比较集中的数据费用高。数据收集的费用在整个调查费用中所占的比例是如此之高,而互联网调查所需的数据收集的费用又是如此低廉,因此研究者可以只需增加不多的总调查费就可以增加比较多的样本量。更为重要的是,对于样本量既定的调查,我们只需额外增加一点费用,便可做更多次的追踪调查。

数据收集速度快是互联网调查的第二个优于其他调查方法,特别是邮寄式调查和面对面访谈的地方。互联网调查收集数据所需要的时间一般为10~20天,在有些情况下可能还会更短。例如,在一项密西根大学学生做的互联网调查中,完成的调查问卷总数中的30%是第一天收到的,而所有完成的问卷的50%则是在第三天收到的(Couper, Traugott, & Lamias, 2001)。互联网调查的问题是回答率的提高比较费时间。因为我们需要时间将有关信息送达那些尚未完成问卷的人和等待那些行动比较缓慢的被调查人完成问卷。

在线的自填式问卷较之邮寄式调查、电话调查和面对面访谈的问卷有以下几个优点:首先,互联网问卷可以包含比较复杂的跳跃式选项,因为跳跃的指令已经编进问卷,因此可以自动执行——被调查人很容易就能理解这些指令。为了达到这一目的,我们可以采用各种各样的视觉辅助手段,如凸现的指示、下拉式清单、图像、录像剪辑,甚至音频提示等。在一份在线问卷中,我们也可将前面的问题答案编入后面的问题。在使用所有与这些最新发明有关的方法时,我们必须提醒读者的是,研究者在使用这些方法时必须慎之又慎,最大程度的减少潜在的回答偏倚。解决这类问题的方法是在设计时采用"最小公分母"。这就是说,在设计采用这些最新技术时,必须考虑那些没有最新的计算机设备和上网速度比较慢的被调查人。如果一份问卷过于复杂,以至下载它的时间过长,那么被调查人便有可能放弃不填,或发回的是只填写了其中一部分的问卷。

有证据表明,能得到相对完整和详细的开放性问题的答案,可能是互联网调查胜过邮寄式问卷调查的又一个优点。但我们只有相当有限的证据可以证明,电子邮件调查得到的开放式问题的答案,其完整程度胜于邮寄式调查(Schaefer & Dillman, 1998)。不过我们不清楚造成这一现象的原因究竟是被调查人文化程度,还是数据搜集的方法。

在人机互动过程中,社会化界面理论(social interface theory)越来越引起研究者注意。在该理论看来,"计算机界面中的拟人化提示可以使用户的反应更类似人与人的互动"(Tourangeau, Couper, & Stieger, 2001)。如果情况果真如此,那么在互联网调查中的拟人化的提示既会带来优点,也会带来缺点。优点可能是虚拟的被调查人(能说话的头像)或其他拟人化技术可以建立比

较友好的氛围，从而使由被调查人自我掌握的调查的结果更为令人满意。但在另一方面，我们也有理由认为，在处理敏感问题或令人难堪的问题，如性行为、酗酒和毒品、投票行为和参加宗教活动时，拟人化的提示有可能产生社会愿望效应（social desirability effects）。令人欣慰的是，近来探讨人性化界面的特性和拟人化技术对互联网调查的影响的研究发现，"无论拟人化程度还是互动的程度都不会对敏感问题产生太多的影响"（Tourangeau et al., 2001, p. 7）。

互联网调查的缺点

到目前为止，互联网调查的最大不足在于，相当大一部分的美国成年人没有上网的途径。据2001年9月的"当前人口调查"估计，56.5%的美国住户有计算机，而能够上网的住户只有50.5%（FCC, 2003）。尽管互联网的使用发展迅猛，但是要能与美国电话的覆盖程度比肩，仍然需要很多年。还有一些专家则对这一天究竟是否真的会到来心存疑虑。正因为如此，我们无法通过互联网设计和执行一个针对全体人口的概率抽样调查，尽管商业公司的报告认为这是可能的（Huggins & Eyerman, 2001）。由于缺乏优质的抽样框，即使对那些可以上网的人来讲，情况也同样如此，这一点也不能不说是互联网调查的一个不足。由研究者主持实施的以概率为依据的互联网调查只限于有限的调查对象，都是那些有着比较完整和精确的抽样框的特定总体，如大学的学生和教师、联邦政府工作人员、某些公司或社团或某些专业组织的成员。

低回答率和由此导致的潜在的回答偏倚则是互联网调查的又一个较大的缺点。通过互联网进行的概率抽样调查不仅得到的信息不多，而且它的回答率一般也都低于邮寄式调查。例如，一个互联网的大学生调查，其回答率只有41%（如果包括不完整的问卷，则为47%），这一数字与研究者描述的其他类似的互联网调查相当（Couper et al., 2001）。在一项有关好客问题的研究中，那些设计将这一研究用于对邮寄、传真和基于互联网调查方法进行比较的教授们，如科巴诺鲁、沃德和莫里奥（Cobanoglu, Warde, & Moreo, 2001），在互联网调查的回答率为44%。有趣的是，这一研究得到的结果与那些认为互联网调查的回答率一般都比较低的观点相左，因为在这一研究中，邮寄式调查的回答率只有26%。但我们对此仍持一定的异议，我们比较同意这样的观点："……为了将互联网调查的回答率提高到类似总体的邮寄式调查的水平，需要做的工作还有很多"（Couper, 2000）。当然，我们应该专注于创造一些能提高在互联网调查中激发被调查人参与热情的方法以提高调查的回答率。在这一方面，邮寄式调查已经有比较成功的经验。这些方法包括人性化的封面、催促提醒的信函和采用一定的奖励等。克服各种可能导致低回答率的困难（如缓慢的通讯速度、不够可靠的互联网连接和不够先进的浏览器）将是互联网

调查面临的一个更为严峻的挑战。

互联网调查会产生特有的回答偏倚。尽管这方面的研究十分有限，但我们仍然有理由认为那些文化程度较低、使用计算机经验较少和计算机设备比较陈旧及使用比较落后的浏览器的人，比那些文化程度较高、有较多使用计算机经验、装备了比较先进的计算机设备和使用更先进的浏览器的人在线完成问卷调查的可能更小。在教育、计算机使用水平和计算机设备的质量之外，有研究者发现，少数民族学生的回答率低于白人学生（Couper et al.，2001）。研究者指出，他们无法确定这一差异究竟来自调查的题目[肯定行动（affirmative action）]，还是来自两个群体之间在对计算机使用和精通程度上存在的差别。显然，这是一个对未来的研究十分重要的领域。

互联网调查与邮寄式调查有某些相似的特点，这使得它们比由调查员进行的调查的效率要低。第一，在线的问卷必须比较短。为了避免过高的无回答率、题项空缺和半途而废，那些曾经使用过各种长度的问卷的研究者发现，为了便于完成调查，问卷长度最多不能超过 15 分钟，而更为理想的长度是不超过 10～12 分钟（Couper，2001）。第二，与邮寄式调查相似，在线的问卷必须是完全不需加以说明的，因为在互联网上是没有调查员可以对含混不清的词义、复杂的指示或问题做出任何解释的。第三，恰如邮寄式问卷显示的形式和内容必须对所有的被调查人都一样那样，在线的问卷所显示的形式也必须对所有的被调查人都一样。但与邮寄式问卷不同的是，在线问卷并不像邮寄式问卷那样很容易就设计成对所有的被调查人显示的形式都一样。问卷设计者必须了解那些影响在线问卷的形式的因素——计算机显示器的尺寸和分辨率、操作系统、浏览器、传输速度等——尽可能地缩小因这些因素在问卷显示形式上引起的差异。第四，尽管在互联网调查中，我们可以对问卷进行设计，以使被调查人按呈现的次序回答问题，但实际经验告诉我们，这种做法是很不恰当的。如果我们要求被调查人在回答下一个问题之前必须先回答前面的每一个问题，那么那些遇到了不愿回答的问题，或无法理解和回答的问题的被调查人就很有可能因此而退出调查，从而导致无回答率的上升。是迫使被调查人在回答后面的问题前回答前面的每一个问题，还是允许他们选择“不知道”或“拒绝回答”这类答案，给他们提供了与其他类型的调查不同的选择（Dillman，2000）。正因为如此，在一个设计精良的互联网调查中，研究者并不对被调查人回答问题的顺序加以控制。最后，无论是做邮寄式调查，还是做互联网调查，研究者实际上都无法对问卷实际上是由谁填写的，或填写问卷时的实际情景（如被调查人是在家，还是在单位、网吧或其他公共场所填写问卷；在被调查人回答问题时，是否有其他人在场等）加以控制。

有些初步研究发现互联网调查与其他类型的自填式调查不同，也许不太适合用于敏感问题的数据收集。其原因可能是因为被调查人担心互联网的安

全性不足以保证自填式问卷的匿名性(Couper, 2000; Dillman, 2000; Cho & LaRose, 1999)。诸如这样的担忧可能导致敏感问题调查的高无回答率、敏感问题题项的高无回答率、报告的不可靠性或三种问题的同时出现。为了确定在被调查人眼里究竟什么样的问题才是敏感问题,以及他们究竟如何处理这样的问题,我们还需要做更多的研究。下面的网址将会提供有关互联网调查的理论和方法发展的最新资料:http://www.websm.org。

电话调查

在一个电话调查中,选择电话号码的方法是多种多样的。电话号码可以随机地选自一本电话簿,或根据现有的电话号码去掉电话号码最后四位数中的一位或多位,再代之以随机方法产生的随机数码。被调查人可以从一份清单中选取,例如从一张含有电话号码的成员名单中选取。我们需要对调查员进行如何与被调查人联系和管理调查问卷的培训(Guensel, Berckmans, &Cannell, 1983)。通常调查员需要根据分配给他们的被调查人或电话号码进行电话调查。此外我们还必须对调查员进行如何用统一和首尾一致的方法收集信息的培训。调查员必须以与书面表达的文字完全一致的问题提问,且对所有的被调查人提问的顺序也必须相同。数据的信度在一定程度上取决于是否遵循诸如这样的标准化程序。自 20 世纪 90 年代中期以来,电话调查遇到了若干挑战,我们将在下文中对这些问题一一予以介绍。

电话调查的优点

电话调查是当今使用最为广泛的抽样调查方法。随着这种方法的若干限制因素被克服,它的使用在过去的数十年间一直不断增长。电话调查被广泛采用的原因之一是有电话住户的增长。1960 年,80% 的住户有电话,而在 1970 年、1980 年、1990 年和 2001 年,这一数字分别达到了 83%,92%,94.8% 和 95.5% (Frey, 1989; FCC, 2003)。因此它的覆盖误差(没有电话的住户)并不大。

研究者担心的另一个问题是没有登录的电话号码的住户数。这一数字因社区而异。但在这些住户所占的比例比较大的时候,把他们排除在调查之外则可能会产生比较严重的问题。调查抽样公司(Survey Sampling, Inc.)(Piekarski, 1997)估计在某些大都市地区,未登录电话号码的住户数量不多,例如在佛罗里达的巴拿马城这一数字为 6.4%,而在威斯康星的希博伊根和奥克拉荷马的伊尼德,这一数字分别为 7.8% 和 13.4%。在其他地方,有相当大一部分的住户没有登录他们的电话号码,例如在德克萨斯的圣安东尼奥和

加利福尼亚的旧金山这一数字高达 38.4% 和 64.4%。大都市地区中没有登录电话号码的住户比例最高的城市是加利福尼亚的斯托克—洛代，为73.4%，实际上最高的 24 个没有登录电话号码的住户的大都市地区，有 23 个都在加利福尼亚。阿拉斯加的安克雷奇是唯一的位于加利福尼亚之外的大都市地区，它的这一数字为 67.6%。虽然随机数码拨号(RDD)已经使研究者克服了这样的问题，但是它的费用要比通过电话簿选取被调查人的费用要稍微高一些。

研究者对电话调查收集的数据的质量也很关心。某些研究表明电话调查的数据质量与其他调查方法大致相当，但对所需收集的数据形式和数量必须有所限制。例如，如果被调查人必须查阅有关记录，或调查时间大大超过 35 分钟，那么电话调查可能不是我们应该使用的最佳方法。不过，如果调查员询问的问题大多数是不需查阅记录的，而是来自记忆的有关态度和行为的问题，且调查的时间在 30 分钟之内，那么电话调查便是一种比较可取的方法。

电话调查的费用居于邮寄式调查和面对面访谈之间。对一个电话调查来讲，我们需要具备以下条件：

- 调查员。
- 一个相对集中的电话中心和调查员工作的办公场所。
- 如果条件允许，提供一定的监控设备。
- 构建或购买总体的样本框。
- 一定数量的工作人员。以对调查结果进行监控；对完成的问卷进行编辑、编码和建立分析所需的计算机数据文件；用 CATI 系统软件对问卷、样本信息进行编程、撰写数据文件和清理类别等各种工作。

调查员是研究者的基层代理：研究者都聘用其他人来做调查，而不是亲自去做调查。这样的做法既有优点也有缺点。一方面，这样做的速度较研究者独自一人进行调查要快得多；而另一方面，调查员必须进行挑选、聘用、培训和督导。我们必须确保调查员能逐字逐句地宣读问题，正确无误地给答案编码，并不以任何方式误导调查。我们也必须付给调查员一定的酬劳。电话调查和面对面访谈都是人工密集式的调查。正因为如此，它们的费用要大大高于邮寄式调查和互联网调查。电话调查有几个非常明显的优点。因为据当前人口调查(CPS)估计，美国住户的 95.5% 都有电话，所以 RDD 调查的样本框偏倚很小(但是，农村地区、南部某些地区和经济萧条的地区最有可能是住户电话百分比比较低的地区。只有移动电话的住户的比例在不断增长，这样的住户并未包括在许多抽样框中。但 95.5% 这一数字可能包括了只有移动电话的住户)。而许多被调查人也比较乐于接受电话调查，这可能是电话调查的回答偏倚比较低的另一个原因。然而，有些操外语的人群，可能不是很乐于在电话中接受调查。而为了调查的顺利，我们必须使调查员和被调查人的民族属

性相配，从而使调查员可以说被调查人说的语言。

在对难以联系的被调查人反复进行联系的情况下，电话调查的回答率通常可达到40% ~80%。我们认为，通常情况下，对那些难以联系的被调查人在不同的日子反复联系6~15次为宜。电话调查的回答率在通常情况下都优于邮寄式调查和互联网调查，其主要原因在于，调查员本人不仅可以直接将调查的重要性告知被调查人，而且还可以将调查安排在被调查人比较方便的时候。不仅如此，调查员还可以亲口宣读问题和亲自记录被调查人的答案。故而，调查员的使用在很大程度上减少了被调查人的负担。

调查地点也会对回答率有所影响：与调查的城市的大小恰好相反，小镇和小城的回答率通常高于大都市地区。

电话调查的另一个优点是数据收集的时间与其他的大多数调查方法相似或更短。打电话是一种与个人或住户开始建立联系的比较快捷的途径，而反复的联系则是与比较难以联系的人建立联系的有效手段。

电话调查样本的地理分布可以比较广阔，因为购买所有地区代码和美国电话的前置号码的抽样框，或以清单辅助的样本是比较容易和便宜的。然而，全州或全美调查的费用却略微高于地方城市的调查，因为这时不仅购买和/或选取样本的费用较高，而且我们还必须支付长途电话费。但地方城市调查花费的时间却几乎与全州或全美的基本上一样。它们之间的主要差别在于，全美调查项目的调查员上班时间要比地方城市或全州调查的更长，以使调查的时间能覆盖全美所有的时区。

调查员的使用带来的一些优点，提高了问卷的功效和数据的质量。由于使用训练有素的调查员，使记录的答案的质量达到相当的高度。此外，因为调查员做了所有的工作——宣读问题、记录答案——如果调查的题目比较有趣的话，那么调查有可能进行30分钟以上。另外，因为调查员受到了如何和以什么顺序对每一个问题进行提问的训练，问卷设计或编排可以比较复杂。例如，我们可以采用多重跳跃的问卷，使被调查人在回答某些问题时可以其他问题的答案为根据，或略而不答。训练有素的调查员可以应对各种情况和处理各种各样的问卷格式，例如他们可以对被调查人回答不够清楚的问题做进一步的追问。

使用调查员的另一个优点是他们能够对问题的顺序加以控制。接受电话调查的被调查人不知道接着会问他们什么问题或调查的问题的数目。如果他们看到了要调查的问题有多少，或了解到如果对某一问题的答案选择了"是"，便要接着回答后面的一连串问题时，他们就很有可能因此而不接受调查，或对第一个问题回答为"否"。

最后，调查员有可能通过电话与被调查人建立起融洽的关系，从而使他们接受调查，认识研究的重要性和意义，进而提供完整和精确的答案，即使对比

较敏感的问题也能如此。在使被调查人自愿回答而没有胁迫感的问题上，虽然电话调查与其他调查方法不相上下，但在进行敏感问题的调查，如可能导致某些疾病的性行为(Czaja, 1987—1988)、后天免疫系统综合症(艾滋病)和高危行为及其他类似问题的调查时，电话调查也不失为一种很好的调查方法(Binson, Murphy, & Keer, 1987; Catania et al. , 1996)。

电话调查的缺点

电话调查的确存在某些不足之处。诸如移动电话、来电显示、寻呼机、传真机、互联网通路、调制解调器、应答机、阻截和呼叫转移这样的电讯技术使地区代码和电话号码在 20 世纪 90 年代成倍地增长，使得与一个住户及其居住者的联系变得更加困难和费时。与此同时，推销的电话变得越来越令人心烦，以至联邦贸易委员会(the Federal Trade Commission)不得不推行全美不接受电话推销登记(a National Do Not Call Registry)。虽然电话调查者与慈善机构、政治组织和被调查人从业的公司都免于受注册登记的限制，但是注册登记无疑是一种清楚的信号。它表明，一般公众对自己找上门来的电话的不满在日益增长。由于这些技术的发展，为了完成一次调查需要进行试拨电话的平均数一直在增长，与此同时，调查的回答率却一直在下降。

电话调查的问题必须简短，且给被调查人宣读的、供他们选择的答案不仅不能太多，而且也必须简短。否则的话，被调查人就无法记住全部的信息。每一个句子最好在 20 个单词以下，语言要简单明了。每一个问题的答案类别以 4 ~ 5 个为宜(Payne, 1951)。在问题的句子太长或答案的类别太多时，被调查人可能只能记住最前面或最后面的那些内容。我们将这些现象称为首位或近因效应(primacy or recency offects)(Schuman & Presser, 1981)。但有一个值得我们注意的例外，那就是苏德曼和布拉德伯恩(Sudman & Bradburn, 1982)认为，在询问敏感行为的时候，最好能使用长而啰唆的问题。

不能提供可视手段，如图片、产品样品，或多项选择题可供选择的答案清单，也是电话调查的一个缺点。为了克服这一缺点，我们尝试过各种不同的方法，但收效甚微。例如，如果我们有被调查人的姓名和地址，我们可以事先寄一封信给他们。这种方法类似一种邮寄式调查和电话调查的组合。在使用这一方法时，我们必须在电话调查进行之前，发一封封面信件，说明有关情况。通过购买那些登录了电话号码的样本被调查人的姓名和地址信息，用 RDD 调查进行组合调查或许还是可行的。不过对 RDD 样本中登录了电话号码的那部分被调查人事先用信件联系的做法已经产生了混合效果(Parsons, Owens, & Skogan, 2002; Singer, Van Hoewyk, & Maher, 2000)。

卡宁、查普曼和哈尼曼(Kanninen, Chapman, & Hanemann ,1992)在一项调查中采用的顺序是电话—邮件—电话，但最终得到的回答率是很低的。该

调查询问被调查人他们愿为有可能在加利福尼亚的圣华金河实施的5个环境保护项目支付多少钱。先依据电话号码的RDD样本一一拨号,然后确定合格的调查户。在最初的拨叫过程中,调查员先向被调查人解释了调查的目的,然后要求他们提供自己的姓名和住址,以便邮寄一份简短的问卷给他们。访谈通过电话进行,而邮寄问卷的目的在于帮助被调查人理解问题的顺序和题意。在最初的电话联系过程中,调查员也要求被调查人确定合适的调查时间。三分之一以上的被调查人拒绝告诉调查员自己的姓名和地址,因此降低了最终的回答率,从而可能使回答偏倚有所上升。大多数拒访都发生在调查员恳请被调查人提供姓名和地址的阶段,因此我们不提倡大家采用这种方法。

在调查过程中调查员无法控制回答问题的情景以及被调查人难以查阅自己户内的有关记录是电话调查的又一个不足之处。在对被调查人进行调查的时候,我们永远也无法知道电话在哪里,房间里是否有其他人在场,如果有其他人在场,被调查人是否感觉不自在。在用电话进行调查时,如果没有事先通知被调查人,被调查人也会难以查阅有关记录。

电话调查的另一个不足之处是从开放性问题那里得到的答案十分有限,这一不足可能不那么明显。电话调查似乎能像面对面访谈那样深究被调查人对问题的看法,并逐字逐句地对答案加以记录,但是实际上被调查人给出的答案只是一句话或几句很短的话。如果调查员试图进一步探究"其他的原因",要求被调查人做进一步的解释或澄清,换来的常常是被调查人的沉默或"死时间(dead time)"。正因为被调查人不愿说得更多,他们便会对追问感到不安或焦躁。有关研究表明,电话调查得到的长而详细的答案远不如面对面调查那么常见(Groves & Kahn, 1979)。

最后,电话调查得到的有关拒访和无法联系的信息相当有限,除非抽样框中包含了被调查人的姓名和地址或其他可以确认的信息。如果我们有被调查人的姓名和地址,那么我们就可以在这方面有所作为,例如比较不同性别、城市的不同地区或其他特征的不同的回答率的差异。在进行RDD调查时,这样的比较只有部分的可能性,除非我们已通过电话进行了联系,或做完了一个调查,否则我们不能确定一个电话号码必定是一个住户的电话号码,而确定哪些人生活在住户内的可能性还要比这小得多。

面对面调查

面对面调查也是一种个人访谈调查。这种调查通常由调查员在被调查人家中或调查人认为比较方便的其他地点进行。其最主要的特点是被调查人和调查员同在一个地方。

这是四种调查方法中最贵的一种，因为它需要很多的交通费和时间收集数据。调查员的报酬大概只占总费用的 25% ~40%，而交通费、答案编辑费和其他费用，则占了 60% ~75%（Sudman, 1967）。为了了解调查路途究竟需要多少时间，我们假定需要调查 800 个生活在芝加哥这样的大城市的住户中的、年龄在 18 岁及以上的人。我们还假定每一选出的街区平均调查 5 个人。如果我们要求每一个调查员平均调查 25 个人，那么一共需要 32 个调查员（800 ÷ 25 = 32）。每个调查员将被分配到 5 个街区（25 个被调查对象 ÷ 每个街区 5 个调查对象 = 5 个街区）。芝加哥市的面积大约为 234 平方英里，有超过 10 000 个的街区。因为街区是随机抽取的，每个调查员到达分配的街区路程可能是 1 英里、5 英里，甚至更多，而为了联系选出的被调查人则可能往返多次。不言而喻，路途花费的时间成为调查所需花费的总时间中的主要部分。

面对面调查的费用明显高于只需花费邮资的邮寄式调查，或"让手指行走的"、在电话中"旅行"的电话调查，那么研究者为什么还是希望进行面对面的调查呢？诚如我们已经简单地讨论过的那样，这种方法之所以是某些调查首选的收集数据的方法，乃是因为它不仅非常适合于某些形式的问卷题项，且能显著提高数据质量。尤其是在使用 CAPI 时，情况尤其如此。

面对面调查的优点

尽管面对面的访谈是最贵的调查方法，然而费用过高的不足却可为它的许多长处所弥补。这种调查方式的费用之所以高于其他调查方法是因为调查员必须前往被调查人的家中，而支付的费用一般既要包括他们路途花费的时间，也要包括他们调查所花费的时间。下面我们以社会综合调查（General Social Survey，GSS）为例，对如何进行一个全国性调查的抽样和访谈做一个简单的介绍，进而说明这一方法为什么费用很高。

假定我们打算做一个全美 50 个州的、年龄在 18 岁及以上的成年人的调查。在翻阅了现存的各种清单之后，我们确定没有任何清单或抽样框含有美国 18 岁及以上的居民的姓名和住址。例如，州机动车管理部门只有持有驾驶执照的居民的信息；选民登记表则不够完整，因为并非所有选民都进行了登记；所得税登记表不仅是严格保密的，而且并非每一个人都完整地填写每年的申报表；即使是美国的邮电服务业，也没有一份完整的用户住址清单。

尽管不存在一份美国居民或住户的完整清单，其完整程度足以作为我们的抽样框，但我们可以使用美国 2000 年的人口和住户普查的数据，确定各种行政或地理区域内的人口和住户数。例如，我们可以相当精确地了解 2000 年 4 月居住在芝加哥市或威克县、北卡罗莱那和马里兰的学院公园的人口数。

GSS 的设计仿佛一个漏斗。首先在使用普查信息阶段，随机地选取比较

大的地区。在 GSS 调查中,在第一阶段做出的选择有 100 个之多。在第二阶段选取的是第一阶段选取的地区的子类或部分。在市内,我们可以使用街区群或普查单位(人口在 2 500 ~8 000 的地区);在县内,我们可以使用城、镇或街区,而在农村地区我们可以选取地块。在第二阶段我们的选择数为 384 个,其根据是第一阶段选取的人口和住户单位。第三阶段通常要选取的是部分街区、列举的行政区或地域。在 GSS 调查的这一阶段,选取的单位最低必须含有 50 个住户单位[5]。这三个最初的抽样阶段所依据的只是有关地区内的人口或住户单位数。在第三阶段之后,"制表人"将被派往每一选出的地区,他们的主要任务是列出这些选出的地区的每一住户单位的清单。在列出了所有住户的清单之后,我们将要抽取住户的随机样本[6]。在住户的随机样本中的每一户,调查员将随机地抽取一位年龄在 18 岁或以上的成年人作为被调查人。平均每一街区或地区需要 2.7 个调查员(Davis, Smith, & Marsden, 2001)。你了解每一街区或地区平均需要 2.7 个调查员这一数字是如何得到的吗[7]?

面对面调查在抽样和数据质量方面有几个优点。它的回答率一般都高于电话调查。其原因之一是在正式进行调查之前,我们可以预先发一封信函到被调查的住户。这一信件的抬头通常印上调查资助者的名称,并对调查的原因、对每一样本户进行的调查的重要性、调查数据的使用方法和对被调查人提供的答案予以保密等做出解释。这封信不仅使调查合法化,而且还能使取得被调查人合作上变得较为容易。另一个与回答率有关的优点是,被调查人比较难以在与调查员面对面的情况下拒绝接受调查。然而,面对面访谈也存在某些不足之处。在犯罪率比较高的地区,进行面对面访谈存在一定的安全问题。在公寓大楼中进行访谈,同样也是一个问题,因为调查员往往很难进门。与电话调查一样,小城镇的被调查人合作程度比较高,而在大都市地区则较低。

面对面调查法的抽样框偏倚一般较低。在用普查数据作为抽样框时,总体中的所有个体,在理论上讲,都有被选入样本的机会。但是随着时间的推移,数据开始过时,普查数据便成为了一个问题。这一问题在人口高度上升或下降的地区尤其严重。在高度上升或下降的地区,研究者必须用地方政府的统计数字;在这一地区近年来进行的调查,或诸如当地房地产经纪人、银行提供的有关信息,来更新人口或住户单位的数字。

面对面调查法的回答偏倚一般也比较低。所有类型的被调查人在面对面调查中大致相当。因为数据收集工作由调查员完成,因此面对面调查有与电话调查相似的许多优点。但是面对面调查能比电话调查更好地控制调查现场。例如,若调查员确信房间中其他人的存在有可能对调查产生影响,他就可以要求将调查换到其他房间进行,以保证能与被调查人单独相处,不受干扰。在面对面访谈中,调查员和被调查人之间的关系也更为融洽,因为被调查人可

以亲眼目睹向他提问的人。面对面调查记录的答案,质量堪称上乘,因为调查员都接受过怎样提问和记录答案的良好训练。此外,他们的工作还将接受督导和编码员的双重检查。然而,面对面调查也存在潜在的来自调查员意想不到的影响。与电话调查和其他调查方法不同,面对面调查调查员的相貌、行为举止和面部表情全部都暴露无遗,因而有可能对被调查人有所影响。

面对面调查法的许多优点都与问卷本身有关。面对面调查的问卷可以比较复杂,因为问卷调查由训练有素的调查员进行。再加之调查员和被调查人同处一室,因而面对面访谈可以完成比较复杂的调查任务或询问比较复杂的问题。例如,我们有一张危险程度各不相同的 25 种行为的清单,这些行为包括蹦极跳(bungee jumping)、延缓张伞跳伞(sky diving)、骑山地自行车(riding a mountain bike)等。这 25 种行为分别单独列在 25 张卡片上,我们希望被调查人把这些卡片分成四堆,每堆分别代表"永远不会做"、"可能考虑做"、"至少已经做了一次"和"已经做了一次以上"四种行为类型。如果调查员和被调查人在同一个房间内,并当面向他解释和演示如何操作,那么这样的工作就很容易完成。面对面访谈也可以使用可视性辅助手段。对于比较长的问题和答案类别,调查员可以递给被调查人一页印有问题或答案类别的纸或卡片,并在自己宣读问题时要求被调查人随着宣读的进度阅读问卷。从而使他们能比较容易地了解所问的问题是什么。在面对面访谈中,调查员也可以对提问的顺序加以控制。

面对面调查法同时也是最适宜使用开放式问题的调查方法,因为这样的调查气氛和节奏比电话调查更为放松。正因为如此,调查员比较容易追问一些额外的信息,而被调查人也不会因为在各种答案之间做较长的停顿而感到不自在,因为他们可以亲眼看到调查员在做什么(Groves & Kahn, 1979)。

面对面调查的时间可以比电话调查更长,其原因有以下几个:它在被调查人家中进行;被调查人在接受调查时不需要总是拿着电话,"死时间"和较长的停顿不再是一个问题,而调查的问题和希冀完成的任务可以更多样、更长和更有趣。面对面访谈法的另一个优点是被调查人可以查阅家中的有关记录。尽管如此,因为调查在被调查人的家中进行。在调查员在场的情况下,他们仍然可能不便马上去寻找有关资料,除非他们事先得到了通知,了解即将进行的调查可能需要查阅有关记录。正因为如此,邮寄式调查和互联网调查也许在这一方面更具优势,因为被调查人可以在他们方便的时候寻找和查阅有关资料。然而,在用 CAPI 做同组追踪调查的时候,来自以前的调查信息可以用程序编入问卷,用作对被调查人的答案进行检查或帮助他们进行回忆的依据(Nicholls II, Baker & Martin, 1997)。最后,面对面调查法也与其他调查一样,宜于询问无胁迫性的问题。

面对面调查的缺点

面对面调查法的缺点主要有四个。首先是费用。诚如前述,完整地登记住户单位和支付调查员的路途时间和其他费用是非常昂贵的。一个全国性的、访谈用时在45~60分钟的面对面调查,每完成一个访谈的费用可能至少要250美元。尽管我们手头还没有可以比较这四种调查方法的费用的确切数据,但经验告诉我们,面对面调查的费用约为类似的电话调查的两倍。

面对面调查另一个明显的缺点是耗时过长。格罗夫斯和卡恩(Groves & Kahn, 1979)的研究数据表明,面对面调查整个过程需要的时间是类似的电话调查的2.6倍。考虑到调查可能遍及全国各个地区所需的路途时间、培训的时间、完成的问卷和材料来回传递的时间及其他各种后勤支援所需的时间,面对面调查数据收集阶段的时间,轻而易举地就可能是类似的电话调查的3倍。

为了在同样的费用下办更多的事,研究者试图在选取的每一个街区中做一个以上的访谈。在某一选取的被调查人不在家的时候,可以使用抽取的另一户。不过我们有可能因为聚类(clustering)或在每一街区做一个以上的访谈而受到惩罚。因为具有相似特征(如收入、人口统计学特征和态度)的人倾向于居住在同一街区或邻里地区。正因为如此,当我们在每一街区抽取一个以上的样本时,那么我们可能在某些变量上收集到相似的信息。具有相同收入和生活方式的人,其价值观和行为方式也往往相同这一点似乎是不言而喻的。一个被调查人固然不可能与他的邻居完全相同,但是较之那些生活在另一个街区或数英里之外的人无疑会有更多的相似之处。因此,当我们在每一个街区做一个以上的访谈的时候,我们并未抓住样本总体中存在的全部的多样性。然而一个街区中的所有居民在我们收集信息的那些变量上未必是完全相同的。抽样统计数据已经告诉我们,最佳的设计总是含有一定的聚类,而不是完全没有聚类的设计。作为费用和缺乏街区内多样性之间的一个折中,大多数社会科学的调查一般都在每一街区做3~8个访谈。每一街区中在那些重要的变量上相似的人越多,我们在每一街区中访谈的人数就应该越少。

面对面访谈的另一个可能的不足之处是被调查人在回答有关个人行为方式的问题时,往往会犹豫不决。一般讲,数据收集方法越个别化,被调查人回答敏感性问题的可能性就越小。但是有关这方面问题研究的发现似乎是含混的。为数不多的几个研究发现面对面访谈在调查沮丧(Aneshensel, Frerichs, Clark, & Yokopenic, 1982)、性行为(Czaja, 1987—1988)和女性饮酒(Hochstim, 1967)等问题时,得到的回答率很低。另一方面,苏德曼、布拉德博恩和他们的合作者(Sudman, Bradburn & Associates, 1979)发现,在调查宣布破产或因酗酒被捕等问题时,各种调查方法并无差异,而怀斯曼(Wiseman, 1972)则发现在调查流产和计划生育问题时,也不存在差异。近年来,在音频计算机辅

助的自填式问卷调查中(audio Computer-assisted self-administered interviews, ACASI)——被调查人在这种设备的帮助下,通过耳机听取问题,并随计算机屏幕阅读显示的问题,然后键入或选择问题的答案——有关研究表明用这种调查方式调查敏感问题的效果,与用纸和笔的自填式调查基本上相同(Tourangeau et al., 2000)。

面对面调查还有一个不足之处是被调查人在接受面对面访谈时更可能提供社会所愿的答案。这一缺点在进行种族态度研究时表现的尤其明显,研究结果似乎明显受到被调查人和调查员种族特征的影响。坎贝尔(Campbell, 1981)及哈切特和苏曼(Hatchett and Schuman, 1975—1976)发现,在调查员和被调查人来自同一种族时,答案模式是不同的。在种族特征不同时,被调查人更有可能给出社会所愿的答案。苏德曼和布拉德博恩(Sudman and Bradburn, 1982)指出,数据收集方法对调查结果有一定的影响。他们认为在面对面访谈中被调查人更可能过多报告社会所愿的行为,而在邮寄式调查中,这样的可能性最小。

组合式调查方法

没有任何文献规定一个调查只能采用一种调查方法。有时根据研究的问题和对表 3.1 所列的因素所做的评估,最好的方法是若干种方法的组合。

我们用几个简单实例向读者阐述这一问题。假定我们想要进行一项有关大学生的调查。调查的问题包括约会强奸(date rape)、酗酒和毒品等。我们再假定可用于这一研究的资金非常有限。开始我们选择的方法是互联网调查,但是我们很快就否定了,因为许多学生可能不愿意参加,因为他们担心在线问卷调查对他们提供的答案无法保密。对学生进行调查的另一个比较省钱的方法是选取几个班级的学生,对选取的班级所有学生采用自填式问卷进行调查。这样每做一次调查,我们就可以得到整班学生的答案。假定,大多数教授都同意在他们的课堂上占用一定的时间进行调查,那么这样的设计效率是相当高的。然而,在我们调查的当天,那些缺席的学生怎么办? 如果缺席的学生很多,那么回答偏倚就有可能相当高。因此,在我们的调查中必须将这些缺席的学生包括进来,或至少调查这些学生的一个子样本,以便对得到的答案进行加权处理。

因为几乎没有教授会允许我们再次到他的课堂对缺席的学生再次进行调查,所以我们必须对这些学生采用邮寄式调查、电话调查或个别访谈的方法来让他们填写问卷。所有这些方法的费用都比课堂调查的高。因此,我们希望从这些没有调查到的学生中抽取一个子样本。调查的学生数量多少和采用的

方法完全取决于我们的调查日程和手中掌握的资源。最快捷的方法是电话调查。在这样的情况下,我们应该如何决策?

本书的作者之一,采用在本书第2章已经简单提到过的组合式调查方法,对医生进行了全国性的调查。这一调查的本来计划是采用电话调查。资源、时间进度和调查问卷都表明这是一种最佳的选择。在试调查阶段,研究者发现回答率非常低——低于50%——即使我们试图与某些医生联系的次数在15次以上。为了寻找提高回答率的途径,研究者进行了两个来自调查总体的医生的焦点组访谈。焦点组得到的一个重要发现是医生(和其他非常繁忙的职业人士)需要一种与他们的工作日程有相当的兼容性的调查方法。通过访谈,研究者了解到,医生希望采用的调查方法是电话调查或自填式的邮寄的问卷。这种组合式的方法果然奏效。正式调查的回答率比三次试调查高了40%左右。

当需要在一个住户内进行多人调查的时候(例如已婚夫妇),我们常常会采用另一种方式的组合调查,即最初对户中的某一成员进行面对面的访谈,将另一份问卷留下由另一位成员自行填写,然后寄回。给邮寄式调查的被调查对象提供在线填写问卷而非纸质问卷的选择,是组合式调查的又一形式。而库珀(Couper ,2000)则对组合式设计给我们以告诫。他引用了两个例子,一个是普查局进行的公立和私立学校的图书馆媒体中心研究,另一个是1999年的"底特律地区研究"。在这两个研究中,研究者最初都在把问卷寄给了样本学校/个人的同时,给每一调查对象提供了在网上填写问卷的选择。在图书馆媒体中心调查中,1.4%的公立学校和不足1%的私立学校选择了在网上填写问卷。在底特律调查中,这一数字为8.6%。显然,研究者必须在这种做法可能带来的好处与在邮寄式调查中提供互联网填答问卷可能增加的费用二者之间进行权衡。

最后一个有关组合式调查法的例子是互联网—电话调查。假定我们想要调查的对象是大专院校的教师。如果大学的名录上刊登了全体教师的电子邮件地址及办公室和/或家庭电话号码,调查的题目、问卷的长度等都适合互联网调查,那么互联网调查配合对那些不回答互联网调查的对象(或他们的子样本)进行电话调查这样的研究设计便是最有效的研究设计了。

使用组合式调查方法的可能性只受到研究者的想象力的限制。与此同时,我们还必须对什么样的方法最适合调查的题目和由于数据收集模式的不同人们给出的答案也有所不同而产生的可能的测量误差这两个问题细加斟酌,再对组合的方式做出适当的调整。狄尔曼(Dillman, 2000)也给出了有关组合数据收集方法的很多很有用处的例子。

注　释

1　计算机辅助调查的形式有很多种(computer-assisted interviewing,CAI):计算机辅助个别访谈(Computer-assisted personal interviewing,CAPI)、计算机辅助电话访谈(Computer-assisted telephone interviewing,CATI)和计算机辅助自填式问卷(Computer-assisted self-administered questionnaire,CASI)。许多政府和专业的调查组织使用 CAI 互联网调查是自填式电子调查方法的一种。而诸如这样的电子调查也包括电子邮件和交互式声音应答(interactive voice response,IVR),它们也被称为音触式数据输入(touchtone data entry,TDE)。本书对于电子调查问题的讨论仅限于网络调查法,因为有证据表明,它们是使用最为广泛和最有前途的电子调查法。

2　在调查最终部署码的使用和报告标准化的尝试方面,美国舆论调查研究会(American Association for Public Opinion Research)出版了一种免费的文献,名为《标准定义:抽样调查个案码和结果比率的最终部署码》(*Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*)。报告涉及的调查方法包括随机数码拨号(RDD)电话调查、个别访谈和邮寄式调查等。报告讨论的内容包括各种形式的最终完成的调查、联系率和如何报告合格性未知的个案、合格的个案和没有全部完成的调查。详细内容请参见:http://www.aapor.org.

3　美国政府主持的一些调查,特别是面对面调查的实际经验告诉我们,货币形式的奖励方法能在不影响数据质量的前提下,有效地提高回答率(Shettle & Mooney, 1999; Ezzati-Rice, White, Mosher, & Sanchez, 1995)。奖励机制对某些难以联系的群体也同样是有效的,这些群体是"……①经济上比较贫穷;②教育程度较低;③少数民族;④青少年、青年和年轻的成年人;⑤吸毒者;⑥经常被调查的专业或菁英人士;⑦行踪不定的过客或那些因为法律或其他原因不愿被人发现的人"(Kulka, 1996, p. 276)。

4　从邮局得到已经搬家的被调查人的信息是很重要的。如果被调查人因为居住的地理位置而不再是合格的调查对象时,或在随后有可能进行追踪调查时,要求做一张住址更正卡,这会对研究者做出正确判断有所帮助。邮局提供的转移地址要收一定费用。

5　最后的选择阶段包括的单位数要比一个调查使用的更多,因为这些地区将被多个调查使用。在同一地区进行多个调查时,抽样和制作清单的费用可以分摊给多个研究者。每个研究选取的住户单位各不相同。

6　此外,还有一点必须提及。调查员不应该居住在与那些样本街区相同的街区,因为,通常调查员最好不要与被调查人相识。在问及诸如收入这样的私人问题时,被调查人更乐于告诉陌生人,而不是熟人。正因为如此,调查员有可能每次都要走 1 英里或更远才能完成样本户的调查,所以这种形式的面对面调查法是十分昂贵的。

7　每一街区完成的调查数不尽相同,因为不合格的住户单位数(如空置单位)和调查合作的程度不同(如拒绝接受调查、居住者难以联系)。一般讲,在 10 个街区中,最终完成 4 个调查的街区有 1 个,完成 2 个调查的有 4 个,而完成 3 个调查的可能有 5 个——10 个街区最终一共完成的调查数是 27,即平均每个街区完成 2.7 个调查。

第4章

问卷设计:编写问题

Questionnaire Design: Writing the Questions

抽样调查的问卷好比一根导管,日常行为和观点的世界通过它进入研究和分析的世界。正是通过它,我们才得以与我们想要研究的现象建立了联系。本书的第4章到第6章,将讨论如何将我们研究的问题转换成可用于收集研究所需的数据的问卷。在第4章,我们首先简要介绍问卷设计的整个过程。然后便是本章的两个主要内容:问卷设计和问题编写的主要问题。第5章和第6章讨论的主要问题是问卷的编排和问题的检验。

虽然不是全部,但我们的大多数例子都来自两个研究项目:一个来自马里兰州成年人的电话调查,内容是有关他们对街头暴力犯罪问题的态度;另一个来自有关马里兰大学学生对本校学院公园校区当前存在的若干问题的看法的自填式问卷调查[1]。这两项研究不仅为我们提供了对一般人口总体和特定人口总体进行抽样调查使用的问卷的实例,还使我们得以了解用于自填式调查和非自填式调查的工具之间存在的重要差别。虽然电话调查或当面调查与邮寄式调查或互联网调查在设计和执行上存在若干差别,我们也必须了解所有这些不同形式的调查在问卷设计的基本问题上并没有什么差别。我们将从这些基本问题开始关于问卷设计的讨论。有些问题似乎很简单,甚至不起眼,但是我们在学习过后将会明白,它们既不简单,也不寻常。

我们一定不要把问卷设计作为一个孤立的问题来处理,而是要把它和可运用的资源及研究的目的综合加以考虑。设计工作的最终目的在于回答若干我们所关心的问题,进而就我们的工具的设计问题做出若干重要的决策。这些问题包括:

- 被调查人是否能提供我们所需要的信息?
- 我们如何才能帮助他们提供精确的信息?
- 为了提供这些信息,被调查人必须做什么?
- 如何将每一个研究的问题重新编写为一个或多个调查的题项?
- 我们应该将什么样的问卷题项包括进来?
- 所选题项是否测量了我们感兴趣的维度?

问卷设计过程

与调查设计和执行的其他部分一样,合理的问卷构建过程是由一个精心计划的一系列步骤组成的,其中每一步骤都需要特殊的技巧、一定的时间和决策。

专业的调查研究者可以用各种各样的方法逐步进行问卷设计,而一个擅长于问卷设计的研究者甚至可以用非常规的方法进行工具设计,以适应不同项目的需要。例如,可以在整个问卷的结构确定之前就开始编写某些特定的问题,可以从那些我们最感兴趣的问题或特别难以量度的问题开始。如果调查将会涉及的若干主题中的某一主题需要编写更多的新问题,或量度的目标特别困难,我们也可以专注于这一个主题。

对那些在问卷设计方面没有太多经验的人来讲,一种行之有效的方法是按照一个系统的步骤来进行自己的问卷设计工作。我们给读者建议的这一组步骤虽然不能说它所规定的问卷设计中各项工作的特定顺序总是优于其他各种工作顺序,但是至少可以说它是具有比较普遍的适用性的,因为我们认为这一组步骤是比较合理的。这样一组步骤的目的在于确保问卷设计过程不会遗漏任何重要的环节,并在制定调查的日程和预算时给我们以帮助。下面我们便来列出这一过程涉及的各个步骤。

问卷设计的步骤

1. 列出研究的问题。
2. 在每一研究的问题中列出调查问题的题目。
3. 列出所有的次要信息(背景变量等)。
4. 通过上网和文献检索,寻找其他调查用过的问题。
5. 根据数据分析的总体计划确认变量清单。
6. 起草调查简介(或封面信件)。
7. 起草新问题。
8. 对问题的顺序做初步编排。
9. 如果需要,对“创建”的问题进行修正。
10. 用某一个同事对工具草稿做测试。
11. 开始修订工作。
12. 根据需要,不时用同事对修订本进行“检验”。

这些系统步骤在帮助我们澄清调查目的同时,也帮助我们开始对它们进行操作化。不仅如此,这一过程也使我们开始懂得如何在正在编写的新题项、

工具的检验和其他各项任务之间保持平衡。

一个特定的调查项目,无论其是否有正式的研究问题,都至少会有一组感兴趣的题目。我们应该先列出所有这些题目,然后再将每一个题目分解成较小的描述性信息。这样的信息形式可以是调查的问题,也可以不是。就学院的学生调查而言,我们可以先列出那些我们感兴趣的问题,然后再在每一问题下列出我们特别希望发现的事物。例如我们是只想了解学生对学校有关酗酒和吸毒的政策的知晓程度和观点,还是同时也希望了解他们的个人行为;我们希望了解的是他们有关学校当前的教学改革的建议,或仅仅是他们自己对当前的教学系统的经验。先列出这些要研究的问题或主题,然后再列出每一主题的分主题。诸如这样的清单,如下面所示:

大学关于酗酒和吸毒的政策

- 对当前政策的了解情况
- 对当前政策的修改意见
- 被调查人的行为
- 被调查人提供的、自己的朋友的行为
- 被调查人对其他学生行为的看法

教学体系问题

- 对当前的教学体系的了解情况
- 当前教学体系的普遍有效性
- 当前教学体系对被调查人的有效性
- 对当教学前体系的切身经验
- 改革的建议

在上面列出的这样的清单中,我们可以轻而易举地发现调查将能得到什么结果,而什么样的结果是得不到的。

然后,应该考虑我们感兴趣的次要信息。此时要考虑的主要问题是我们是只对整个目标总体感兴趣,还是对总体中的子群体也同样感兴趣,或我们感兴趣的是子群体中的个体,还是对它们进行比较。如果我们对子群体感兴趣,那么除非来自样本框的信息能将被调查人进行分类,否则我们必须通过调查得到这些信息。这些信息包括样本的人口学分布信息,如年龄、性别、教育程度、种族和/或收入等,通常这些信息总是我们所需要的。我们也可以从其他适合我们研究的实质性问题的研究中,得到一些与这些人口特征和其他特征有关的数据(我们也有望从中发现情况果真如我们所想)。就学院学生的研究而言,我们则要考虑什么样的次要信息是有用处的。如果我们希望做子群体的分析,以了解不同特征的学生的行为和观点有什么不同,那么我们必须确定这些特征可能是哪些。

某些更加难以了解的总体的其他特征，可能对我们的分析会有所帮助。例如，在犯罪研究中，了解被调查人或他们的熟人是否受到犯罪行为的侵害这一情况可能会对研究很有帮助。举个例子说，对判决所持的态度有可能与被调查人的想法而不是与大多数人口学特征有关。我们究竟是应该对这些因素还是应该对其他因素感兴趣取决于我们的研究目的和假设。我在这里想要说明的问题是，如果没有收集相关信息，那么某些分析就无法进行。

如果我们希望使用的问题可以从其他研究中得到，那么我们就可以省去很多工作。如果研究的主题关系到某一特定的总体或问题，例如某大学的校园问题，那么就不可能从其他研究中找到有用的问题。然而，对于许多研究主题而言，例如那些与犯罪行为损害和刑法体制问题有关的研究，许多调查，包括正在实施的大型联邦调查都涉及到了。

花一定的时间，就研究的主题在网上搜索一下是否有恰当的可资利用的问题是很值得的。当然，我们不可因为这些问题已经用过，而认为它们是没有问题的，或甚至认为它们是完美无缺的。不加分析地采用"以往多次使用过的"或"有经验的研究者使用过多年的"调查问题是人们常犯的一种错误。因为一个在以往未加批评地一直使用的、不怎么样的问题，并不成为我们也要同样不假思索地使用它的理由。"有经验的研究者"可能在自己从事的学科内很有知识，但是在调查设计方面却可能知之不多。使用调查的问题与使用其他任何产品一样，购买者（或在这个例子中是借用者）必须对之有所了解。

我们希望读者在工具设计阶段，要对如何进行分析的问题予以特别的关注。许多研究者都在自己的工具即将完成之前，对照自己的分析方案，对它进行检查——如果他们已经做了分析方案的话。然而，将一些并无实际分析意义的题项包括在调查中，却将其他某些重要的题项或背景变量排除在调查之外，或虽然包括在了调查中，但询问的方式却不宜最终的分析，诸如这样的现象在调查中也经常出现。不过要做到这一步并不是很困难的。

当然，要事先了解所有的分析几乎是不可能的。但是考虑一下调查对象自身的频数分布，或如何根据人口学特征及其他子群体特征设置分析变量以及怎样在简单的交互表中应用这些变量还是很有用处的。我们不必事先做出十分详尽的分析方案，试图穷尽所有的分析。但我们必需对分析有所考虑，哪怕它只是几个表格的草稿、草图或图表说明，都可能对我们的调查和最终的分析会有所帮助。这样的工作只需要花费很少一点时间就可以完成。

调查的准备工作应该从简介（或封面信件）开始，这一点很重要，因为这样才能得到被调查人的合作（这一问题，我们将在第 5 章详细讨论）。此外，编写简介或封面信件需要对有关调查进行扼要介绍，这也是一个非常有用的步骤，可以迫使我们自己进一步澄清调查的基本目的，并用寥寥数语使可能的被调查人对调查的目的了然于胸，这一点是十分必要的，但是我们也不能操之

过急或画蛇添足。对被调查人做的调查目的简介，要详略适度，不必过于详细，但也不能过于简单。我们提供的信息只要能取得被调查人的合作就足够了。这些步骤将会有条不紊地引导我们先完成问题的起草、编排，然后再完成问卷的构建和检验。这些问题后文都会讨论。

问卷设计的要素

在讨论问卷设计和问题编写时，我们关注的重点是回答的过程(response process)、单个问题的效用(utility of individual questions)及单个问题和整个问卷的结构。为了达到这一目的，我们将继续就问卷如何与调查研究的过程相匹配，研究人员、调查员和被调查人三者在数据收集过程中的关系等问题做一个简单介绍。

在问卷设计过程中，我们必须牢记调查进行的目的和实际环境。苏德曼和布拉德伯恩(Sudman and Bradburn,1982)称之为"有目的地交谈"。施瓦茨(Schwarz, 1996)则告诉我们，调查员都受格瑞斯(Grice)提出的"交谈规范(norms of conversation)"的支配。在构建问卷时，如果问卷将由调查员填写，而非被调查人本人填写的话，我们应该更多地考虑这类人际因素。

在过去的15年或更长一段时间内，认知心理学家和调查方法专家合作，对调查研究的设计问题，特别是对问答过程(question-response process)进行了大量深入的跨学科的研究(Jabine, Straf, Tanur, & Tourangeau, 1984)。尽管这些合作研究涉及的大多数方面都超出了本书讨论的范围，但是他们采用的用于整个调查过程的模型(Tourangeau & Rasinski, 1988; Presser & Blair, 1994)，包括他们规定的被调查人的任务，都可以作为帮助我们在工具设计时做出正确的决策的十分有用的指南。所谓模型，无非是对一个过程或现象是如何运作的一个规范的描述。模型的目的在于描述这一过程的组成部分和这些部分是如何在一起运作的。

在这种调查过程模型中(见图4.1)，研究者构想设计了调查的问题，并确定了它们在分析中的使用方法。问题可以由调查员填答，也可以由被调查人自行填答。而被调查人必须做四件事：理解问题、回忆与之有关的信息、决定给出的答案并向调查员报告自己确定的答案。这些答案将被输入数据库，然后研究者将对结果做分析，在完成整个循环之后，它又回到了原来确定的该题项的分析使用上面。在设计问卷的过程中，我们必须牢记这些参与者所扮演的角色和他们承担的任务以及分析的计划，我们放在问卷中的问题都是为了在分析时使用的。

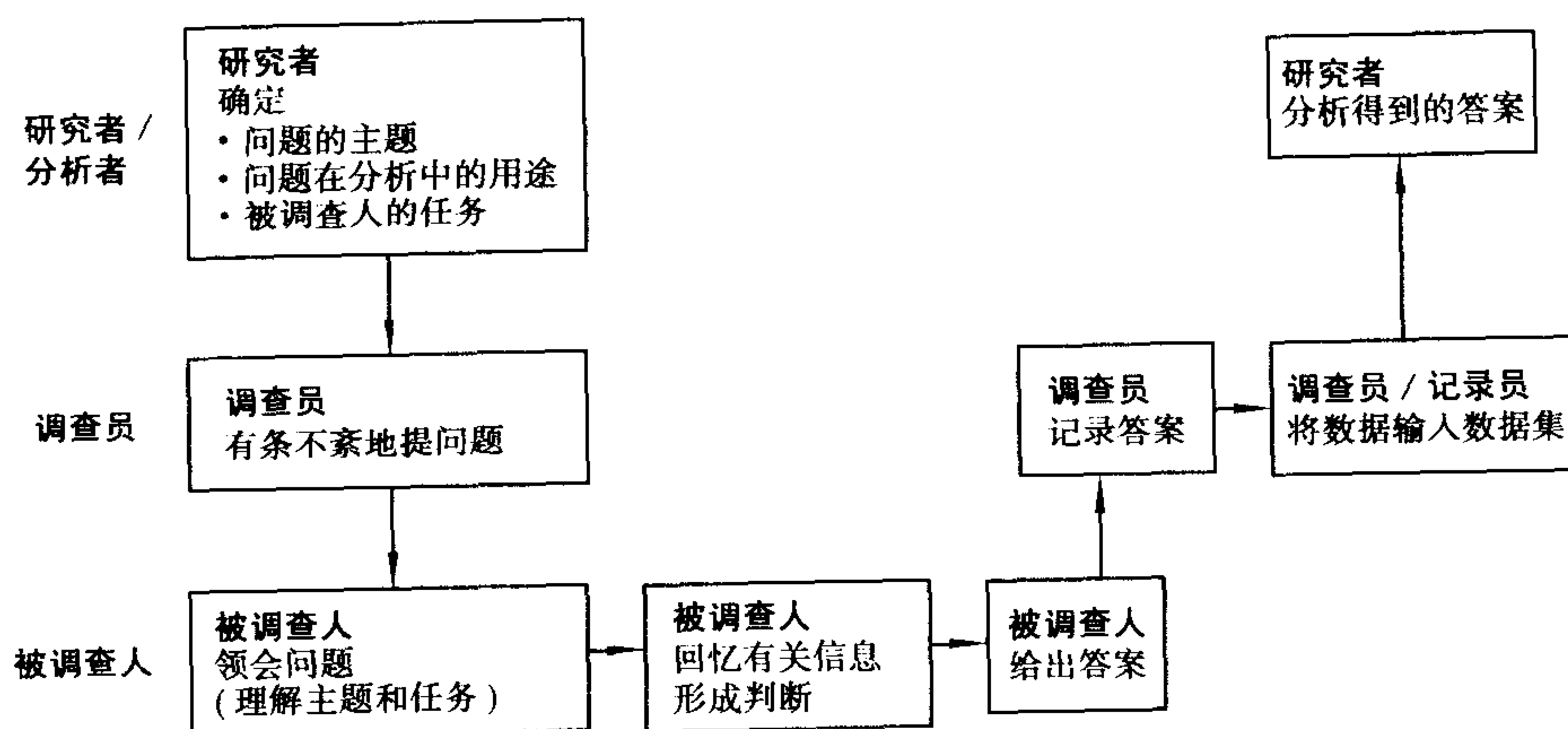


图4.1　调查数据收集过程模型

在参考采用既定的程序、设计技术、以往的研究、经验法则和自己掌握的技巧，开始问卷设计之前，我们必须首先明确自己的目标是什么。明确堪称上佳的问卷的特征是什么？明确我们怎样才能确定自己的目标是否已经达到？调查模型正是我们明确自己目标的指南。

在这一章，我们将会提及一份高质量的问卷所具有的许多特点。但其中三种是最重要的，它们并列于我们列举的各种特点之首：它们是我们感兴趣的因素的有效量度；它们将保证被调查人与我们的合作；它们也将保证被调查人提供我们可以接受的精确的信息。

研究者之所以要确定研究主题和每一个问题在分析中的用途，其目的无非是希望能在调查中得到一个有效的量度，即所谓的测量工具，这一点是不言而喻的。就这一点而言，最重要的是，工具必须能测量态度、行为或研究问题需要的其他属性。每一个研究问题可能需要一个或多个调查题项。但是，如果我们不能用恰当的维度测量我们感兴趣的因素，那么我们所付出的种种努力最终将付之东流。正因为如此，在我们编写（或借用）问题之前，必须首先清楚设定每一题项在分析中的用处。

假定我们问题已经是一个有效的量度，那么随之而来的问题便是它还必须便于被调查人配合调查，并提供我们可以接受的精确的信息。如果我们设计了一份问卷，尽管它很有效，但几乎没有被调查人愿意回答它列出的问题，或虽然愿意回答它列出的问题，但答案却很不可靠，那么我们的设计就是失败的。被调查人不仅必须如研究者所需要的那样来理解问卷列出的问题，还必须有我们所需要的信息以及愿意和能够以问题要求的形式提供答案。

为了达到这一目的，我们必须尽一切努力使我们的假设陈述、句子结构或研究概念的语言表述，不仅能为少数人或专家所理解，而且能为大多数普通的

人所理解。如果我们的问卷既不能在取得被调查人合作方面令人满意,也不能在便于被调查人回答方面令人满意,那么我们要付出的代价便是测量结果的歪曲。当然影响问卷质量的因素并不止三个,但最主要的正是这三个。

假设我们的研究问题是:大多数人是否认为刑事司法系统运转良好? 这时我们不妨设想用一段十分浅显的文字来开始我们的调查设计之旅。我们可以尝试用一个诸如这样的问题作为调查的题项:

您是否认为大多数人都觉得我们的刑事司法系统运行得很好?

1.是　　2.否　　3.不知道/不能肯定

假定通过试调查或只是通过更深入的考察,我们发现这样的题项存在如下的不足:

- 被调查人不清楚人们对刑事司法系统的普遍评价是什么。
- 被调查人对刑事司法系统的理解不尽相同。
- 被调查人反对在"是"和"否"之间做出选择,因为他们的观点在这两种备择答案之间。
- 被调查人希望具体谈谈司法系统的某些问题,而不是做一个一般的判断。

可能对这一问题的批评意见还有更多,但是上述这些意见已经足以使我们归纳出批评意见的几种类型。尽管问题的确是我们研究所感兴趣的问题,但它们的表达方式并不像一个调查题项。现在我们来看一下修改后的版本,它吸收了上述四个意见中的三个。

您认为被判刑的人得到的判决基本上都是公正的、大多数是公正的、有时是公正的,或都是不公正的?

这一题项尽管也存在一些问题——请读者在读完本章之后,作为练习把它们列出来——但是修改后的题项,较之原来的版本的确在三个方面有所改进:它问及的是被调查人自己的观点,而非对他人的观点做判断;它删除了含混不清和过于复杂的刑事司法系统这一专用名词;它提供了一组不同类别的答案,以测量观点上存在的渐变。为了使问题更宜于被调查人的理解和回答,我们采取了各种措施,但在达到这一目标的同时,它却背离了我们研究的问题的本意:大多数人是否认为刑事司法系统运作良好? 为什么我们要这样说呢?

因为问题的新的表达方式至少在三个方面是失败的。首先,它过于狭隘。刑事司法系统的含义远远不止是判决一个方面,它也包括逮捕和释放等其他机制。其次,未能涉及对该系统有所影响的因素,如被告的财力资源,实际案件中,这一因素多少对最后判决的"公正"程度有所影响。第三,这一题项并未反映该系统的全貌,而系统各个部分的运行情况也不尽相同,有些部分好一些,有些部分可能差一些。我们研究的这个问题是比较复杂的,因此调查的大

部分内容都要考虑这一复杂性。如果我们要研究的问题不是一种简单的选择,那么我们可以考虑采用下面这种形式的问卷。这种问卷含有一个导言(或称调查简介)、多个题项和开放式的后续题项。

刑事司法系统的三个主要的组成部分是警察、法庭和监狱。我们想知道您对其每一个部分的运作情况有什么看法。

1. 您认为警察的工作做得非常好、好、一般,还是很差?(如果一般或很差)您希望他们能在什么方面有所改进?

2. 您对法庭的工作有什么看法?他们的工作做得非常好、好、一般,还是很差?(如果一般或很差)您希望他们能在什么方面有所改进?

3. 你对监狱的工作有什么看法?他们的工作做得非常好、好、一般,还是很差?(如果一般或很差)您希望他们能在什么方面有所改进?

这一稿既避免了第一稿存在的一些问题,但又没有丢掉我们研究的问题的中心意思。我们对刑事司法系统一词做了定义,以保证所有的被调查人都能在相同的意义上回答这一问题。我们对该词的三个基本含义分别进行询问,如果人们对这三个部分的认识有所不同,那么我们就可以捕获到这些信息。对那些认为系统某一特定部分运行不够正常的人,我们就可进一步探究其原因何在。诸如这样的多重型的问题增加了我们分析的复杂程度,同时也使分析内容更丰富和更真实。

我们在通过操作化将研究的问题变为问卷的题项的时候,必须力求做到具体确切,以正确无误地指出我们希望从被调查人那里得到什么样的信息。而这一任务将迫使我们再一次审视研究问题的某些方面。如果我们能精心地再一次审视我们研究的问题,并认真考虑这样的再审视对我们的分析计划的效应,将对我们整个调查有很大的帮助。在这个决策的关键点上,我们需要考虑以下几个问题:我们究竟需要什么样的数据?我们怎样才能得到它?我们准备怎样使用它?下面我们将转向模型中被调查人所扮演的角色:正确理解问题的含义、回忆有关的信息、思考并给出自己的答案。首先,我们必须懂得,所谓理解必须包括对问题的主题和自己所承担的任务的理解。例如,假设在学院学生研究中,我们希望了解高年级的学生自入学以来上了多少节数学课。我们草拟了一个问题:"你上了多少节大学数学课?"这时被调查人不仅要理解问题的主题,大学数学课,而且也必须了解在回答问题时如何正确地计算(课时)。课时计算看起来似乎很简单,实际却未必。我们不妨来看一下,下列各项是否应计作课时:

1. 正在上,但还没有上完的课。

2. 大学水平的数学课,如微积分,但却是在高中上的。

3. 在大学上的补习课。

4. 转校的学生,在以前的大学上的课。

5. 上学期不及格的学生上的课。

6. 以数学为基础，但由其他系讲授，或并未作为数学课列出来的课，如统计。

7. 旁听，不记学分的课。

8. 一门只分及格和不及格的课。

9. 复习或补习性的课。

在人们对一个名词的理解上可能包含不同的含义的时候，列出这样一张表一般是很有用的。作为一个练习，请大家编写一个包含上述第3,4和8类课程，但不包括其他类课程的问题。

即使像大学数学课这样一个简单的问题，也需要研究者做出若干个决定，而这些决定将对问题最终如何编写和最终得到什么结果产生影响。正因为如此，在提出诸如计数或估计这样的任务时，我们必须认真考虑究竟要被调查人做什么这一问题。如果我们将什么应该计数这一问题留给被调查人自行决定，那么不同的人便会将不同的事物计数进来，从而将测量误差引入我们所得到的数据。

诚如模型所示，被调查人必须首先理解问题，然后才去回忆有关的信息。有相当多的调查问题有赖于被调查人回忆信息的能力，因此我们有必要对此进行比较详细的讨论。我们不妨考虑一下记忆在调查访谈中的作用。我们倾向于认为，事实、事件或行为方面的问题是凭记忆回答的，其实我们还应该意识到观念方面的问题也会受记忆的影响。如果被调查人被问及有关某些问题的观点此前他已经考虑过，那么他要做的事只不过是回忆以前自己对这一问题所做的判断。如果被调查人以前没有考虑过这一问题，或在调查过程中要求他做一些其他的考虑，为了表达自己的观点，他就必须回忆一些与之有关的信息。这些新的信息将被用来形成自己的判断，然后再将它作为答案报告给调查员(Tourangeau & Rasinski, 1988)。

虽然在问卷设计和改进的过程中，我们并不需要考虑不同的被调查人之间存在的记忆力好坏的差别，但是我们必须明白某些回忆任务本来就是比较困难的。被调查人特有的经历、记忆能力和认知力，以及所采用的回忆策略，都会相互作用，共同对他们回忆的信息产生影响。调查中我们究竟可以问什么样的问题，首先取决于我们对于被调查人可以回忆起我们所需要的信息的清晰程度的判断。然后，我们可以通过用问卷做的试调查，对我们的判断加以验证。

现在我们再回到有关学院学生调查的例子，假定我们希望了解学生是否认为在校园里已经有合适的治疗酗酒和吸毒的计划。有些学生会对这一问题做一番思考，或对这一问题有更多的了解，这也许是因为他们对这一计划有所接触。这样，这些学生因为有以往经验形成的判断，所以很容易回答这样的问

题。而那些还没有形成自己观点的被调查人则必须对诸如这样的事做一些思考,譬如他们曾经读过的有关这种计划的资料,或回忆一下他们的熟人的有关经验。

在另一个例子中,如果我们要求学院高年级学生列出他们上一个学期修过的课程,那么他们很快就可以做到。如果我们要求同一批学生列出他们在九年级修过的课程,虽然有些学生做得到,而有些学生做不到,但无论如何,对所有这些学生来讲,这一问题无疑都是比较困难的。而造成这一任务比较困难的原因可能有多种:时光流逝,诸如其他课程这样的干扰性事件;对学院的高年级学生来讲,九年级上过的课程可能无关紧要。如果我们现在要求他们列出上学期的学习成绩,那么我们可以设想,几乎所有的学生都可以做到这一点。当然,这需要花费的时间,可能比回忆上过什么课程稍微多一些。而他们在初中时的学习成绩是什么这一问题,对所有的学生来讲都是一个更加难以回答的问题,因而我们可以预期,得到的答案可能参差不齐。在执行回忆学习成绩的任务时,我们将会遇到另外一个因素——一致性(uniformity)。对那些所有的课程成绩都为 A(或都为 C)的学生来讲,这一任务比较容易完成,因为他们不需要一门一门地去回忆每一门课程的成绩。

不仅被调查人的记忆能力是影响任务完成的一个因素,任务本身的性质同样也是影响它完成的一个因素。在一项研究的设计过程中,调查研究者经常会讨论(在有或没有来自研究的证据的情况下)一些特定的回忆工作对被调查人来讲究竟有多么困难这样一个问题。问卷中应该包含什么问题和问题应该怎样表达,在一定程度上都取决于他们对这一问题的判断。

我们也必须承认,通过一个调查就想解决所有我们感兴趣的问题是不可能的。有时,一项任务对大多数被调查人来讲实在是太困难了,进行这样的调查,其结果不是被调查人不愿配合调查,就是得到的数据很不精确,或二者兼而有之。例如,流行病学家和老年学家都会有兴趣对年龄在 65 岁及以上老人的青年时代的健康史进行详细的调查,但由于许多有关的事件并无记忆价值,加之年代久远,因而有可能难以从被调查人那里收集到有关的数据。

在改进我们的调查问卷的过程中,我们经常会使用诸如我们刚刚介绍过的那样的推理方法。这一方法对我们的讨论的作用,不仅在于它能帮助我们确定把什么问题包含到我们的问卷中来,而且也促使我们以便于被调查人对问题进行回忆的方式来构建它们。考虑如何构建有关数学课程的问题的题项便是这样的例子。在构建这个问题时,我们发现如果只使用单独一个题项,有可能会使被调查人在回忆有关信息时感到困难,于是我们将这一任务进行了分解,改为分题项地询问被调查人的各个学期的课程或不同类型的数学课(如正常的数学课、补习课、及格—不及格的课程)。

我们所做出的决定影响被调查人回答问题的质量和是否愿意配合调查的

态度。而可以肯定的是,如果我们提出的问题大多数被调查人都感到比较困难,那么他们的合作程度就不会太高。

编写问题

现在我们至少对如何将我们研究的问题通过操作化转变为调查的题项有了一个总体的认识,下面我们便来讨论一下如何开始构建一份问卷的问题。我们可以借用其他调查使用过的问题,也可以自己编写问题,两种方法各有所长。借用以前的调查使用的问题的优点是,它使我们得以(或者是需要)将自己的调查结果与以前调查的结果相比较。如果我们使用的问题与以前调查使用的不同,那么这样的比较就很有问题。即使我们的调查结果与以前的研究有所不同,我们也无法确切地说,这一差异究竟有多大,因为我们使用的问题已经发生了变化。此外,来自其他调查的问题本身也许就来自它以前进行过的调查,已经经过了很多次的检验,无需我们再做检验,能节约一定的时间和精力。

有一类的确需要借用的重要问题是筛选性问题(screening questions),这样的题项通常都位于问卷的开头,目的在于确定合格的研究对象,即合格的被调查人。例如,假定我们要寻找一种有特定特征的人,譬如受犯罪行为侵害的成年人,而我们对所有住户中有这类成员的住户的比例的估计是以以前的研究为根据的。在我们的问题中,必定要有一个用以确定可能的调查对象是否是犯罪行为的受害者的问题。如果我们提出的了解受害情况的问题与以前的研究不同,那么我们多半会发现受犯罪行为侵害的情况与以前的研究结果有所不同,这也会影响我们的筛选率和与之相关的费用。这就是说,如果我们希望从被调查人中间筛选出一定比例的人作为我们的访谈对象,而实际上我们经筛选得所到的访谈对象的比例低于预期,那么我们要么为了得到我们期望的样本容量而增加调查费用,要么保持原来的费用而减少个案数。

一般讲,我们必须对借用的问题能否真正测量我们打算测量的变量这一问题持慎重的态度。未经深思熟虑地借用一个问题,似乎节省了我们的时间和精力,实际上它会使我们在结果分析时花费更多的时间和精力。

即使是借用的问题我们也必须进行试调查。语言习惯和它的内涵都会随时间的变化而变化,我们问卷的使用环境也可能和以前的研究有所不同(Converse & Presser,1986)。如果我们操作管理的模式和原来的有所不同,那么借用的题项也必须做试调查。一个在自填式问卷中运作良好的问题,可能在电话调查中听起来不是很清楚,或产生其他的问题。最后,我们必须注意的问题是,我们借用的问题原来所适用的总体。一个适用于医生调查的问题,可能完

全不适用于病人或一般人口总体的调查。尽管我们在这里提出了很多警告，但是我们必须指出，可资利用的高质量的调查问题的确有很多，无视它们的存在是很愚蠢的。然而，考虑到为了使借用的问题能用于我们自己的研究，它们也必须通过同样的检验，我们将在下面对检验标准做一些讨论。

调查问题质量的评价标准

我们必须在资源有限的前提下，设计和实施我们的研究，不可能把我们感兴趣的所有问题都包含在问卷中。每一个问题不仅需要一定的费用，而且也必须是比较恰当的。表 4.1 列出了一个问题在用于研究时所必须满足的几个标准。这一决策指南对我们讨论过的问题进行了总结，并把它们按步骤顺序编排在了一起。我们可将它们应用于考虑将哪些问题引进我们实际的研究。决定问题是否符合标准，依靠研究者的判断。有些调查问题，在我们可以对之做出取舍之前，可能要先做试调查。而另一些问题则不需要，因为问题比较明显。这一组标准可供我们在整个问卷的设计和改进过程中参考。

表 4.1　关键性决策指南：问题的效用

1. 调查的问题是否测量了某一调查的问题的某些方面？
2. 问题提供的信息是否为其他有关变量所需？

（如果 1 和 2 为否，删除。如果其中一个或两个都为是，通过。）

3. 大多数被调查人是否都在同样的意义上理解这一问题？

（如果不是，修正或删除。如果是，通过。）

4. 大多数被调查人是否都有回答这一问题所需要的信息？

（如果没有，删除。如果有，通过。）

5. 大多数被调查人是否都愿意回答这一问题？

（如果不是，删除。如果是，通过。）

6. 分析这一问题是否还需要其他的信息？

（如果不需要，继续。如果需要，且如果有其他信息可资利用或从其他调查中收集，继续。）

7. 这一问题是询问所有的被调查人，还是只问其中的一部分？

（如果是全部，继续。如果只是一部分，且如果这一部分是事先或通过问题可以确定的，继续。）

指南中前两条与问题是否为调查所需要密切相关，后面几条则是关于信息的可获得性。不言而喻，如果一个题项测量了我们研究的某一个问题的某一方面，那么我们就应当考虑把它包含在我们的研究中。如果一个题项本身并不直接与我们研究的问题有关，但对另一题项有补充，那么也应该把它包含在我们的研究中。例如，假定我们第一个问的问题是："自从住到这一地区以

来,您是否遇到过不得不报警的情况?”接着再对那些回答“是”的被调查人追问第二个问题:“报了多少次?”为了进行比较,我们还必须另外再问第三个问题:“您在这个地区已经住了多长时间?”尽管我们对这一问题本身并不感兴趣,但它却是分析第二个问题所需要的。

调查问题的结构

虽然可资研究者利用的问题的形式不胜其多,但大多数调查实际需要的只是不多的几种,而这些实际需要的问题的大多数,其形式并不是很复杂。不过诚如我们将要了解的那样,许多第一眼看来似乎很简单的题项,常常会给被调查人带来严重的问题,需要研究者做出相当困难的决定,同时还有可能耗费许多资源对它进行检验和修订。而这时,问卷设计和改进需要的时间和其他资源常常超出初涉调查研究的新手所能承受的范围。

在编写问题的时候,我们必须考虑到数据收集过程模型中的两个部分——被调查人和调查员。我们提出的任务必须是这二者都能很好地完成的。尽管我们的注意力主要放在了被调查人身上,但是我们必须记住,调查员必须能流利地念出问题、清楚而明白地读出每一个字(尤其是在电话调查中,因为那时调查员不能借助视觉提示手段消除被调查人的疑问),出于这样一些考虑,我们增加了很重要的一点——我们希望被调查人的回答不受任何因素误导。正因为如此,在设计和改进问题的时候,我们需要不断地问自己,数据收集的过程本身是否产生了偏倚。

我们所要编写的问题应当是朴实无华和简单易懂的,恰如一个律师所提的问题那样明白无遗和直截了当。如果问卷由调查员填写,我们希望问题的文字听起来朴实自然;如果问卷由被调查人自行填写,我们希望问题的文字比较流畅,那么很重要的一点就是我们必须认识到,调查问题的构建是一项非常特殊的工作。它之所以特殊是在于它有清晰明确的目的。我们往往需要花很多时间和精力才能使问题的文字自然流畅,因为调查的问题不是一种简单的交谈方式。从理想的角度看,精心设计的调查问题应该去除了日常语言存在的模棱两可、自相矛盾、不得要领和叠床架屋等弊病。再者,尽管我们花费了很大力气以保证我们的问题的确能反映自己的研究目标,但这些问题的文字一般讲并不是任何一种社会科学的语言。这就是说,问题使用的文字比它们所承载的科学目的简单。调查问题的文字既不同于日常语言,也不同于科学论述,它们必须在没有任何补充解释的情况下就能被理解。调查的问题一旦定稿,我们要求调查员每次都必须以完全一样的方法提出,文字既不能增删,也不能更改。尽管我们经常不知道为什么问题的措辞不同会导致结果的不同,但这样的影响却是不容忽视的(Schuman & Presser,1981)。通过调查研究所得到的数据的可靠性,在很大程度上取决于被调查人在填写和理解问题时

的一致性。

在对那些我们编写的、选自草稿的问题进行文字的修饰、排列前后顺序、编制答案的测度和确定供被调查人选择的答案形式时，有几种组合方式可以选择。我们先将所有可能的组合方式罗列出来，再找出大家都认为比较好的那种，然后讨论为什么有的时候还要对这种组合方式做一些必要的变动。

无论是电话调查、面对面调查还是自行填写问卷的调查，我们可以选择的开放式或封闭式问题都是广泛的。对于前者，我们未曾给出供被调查人选择的、确定的答案，而在后者，我们则给出了确定的答案。我们推荐大家尽可能使用封闭式问题，原因后文会说明。由于来自开放式问题的数据是叙述性的，必须进行解释和编码，在调查完成之后，研究者得到的结果仍然不能立即进行量化分析。因此，我们将对开放式问题进行的讨论有限，我们的讨论主要集中在封闭式题项上。

一个封闭式的问题由两个部分组成：陈述的问题和分类的答案。我们必须对每一部分予以同样的关注。我们要尽可能直截了当地陈述问题，同时我们要求给出的答案必须是我们所提供的各类选项中的一个，且一般讲只能是一个。一般情况下，特定的要求或条件都是问题的一部分。这些附加的部分对问题的含义做了与我们研究目的相对应的限定。最为简单的例子是选项为是—否的答案。例如：

在过去的一年，您认为您个人或您的财产是否受到犯罪行为或犯罪企图的侵害？

1. 是　　　　　　　　2. 否

虽然这时我们已经做了某种决定，如决定不包括“不知道”或“不能肯定”，但答案选项的选择还是比较简单明了的。至于是否要在问卷中包括诸如这样的选项，则取决于我们对许多被调查人是否都能回答只包括“是”或“否”两类答案这一问题所做的判断。

设想某一个被调查人有一天回到家的时候，发现一楼的窗户被打破了。这是意味着有人企图破窗而入，还是小孩们在玩耍时不小心打破的？尽管她对这一问题发生的原因搜肠刮肚，但仍然百思不得其解。现在再让我们来设想一种情况：几个月以前，某一个被调查人在单位，稍不留意，她的计算器就被人偷走了。这个计算器所值无几，她觉得已没有希望再找到它，因而并没有报告警察或保险公司。在她听到我们调查的问题时，她不能回忆起这个微不足道的盗窃事件的确切时间。这件事究竟是不是发生在去年？尽管她想了又想，但还是可能无法确定。

对这一问题，我们觉得也许增加“不知道”这一选项比较恰当。但这些例子告诉我们的东西还不止这些。它们告诉我们，我们对每一个问题都应做周密的思考，从变化多端的实际情况到所有或大多数被调查人是否能以我们期

望的方式回答我们提出的问题。通常在考虑这些问题的时候,如果用几个假设的例子,设身处地考虑一下,或把设计的问题草稿请自己的同事试答,便可以很快地找到问题的症结所在。诸如这样的非正式的判断和交谈虽然不适合用于正式的调查,但无疑对它是一种很好的补充。

作为一个练习,请大家考虑一下,在诸如下面这样的问题中,某些被调查人可能会有什么样的困难?

由于担心受到犯罪分子的侵害,您会避免去一些您本来打算去的地方,您避而不去某些地方的情况是:

1. 总是避而不去
2. 经常避而不去
3. 几乎没有遇到这样的情况
4. 从来都没有遇到过这样的情况
5. 不知道(自愿选择)[2]

在过去的一年,您认为在您居住的地区,暴力犯罪的情况有所改善、变得更严重,还是没有什么变化?

1. 有所改善
2. 变得更严重
3. 没有什么变化
4. 不知道(自愿选择)

如何对调查的问题做具体规定

诚如前述,在许多时候,我们都需要对一个问题在一个或多个方面做出规定或限制。为了能使被调查人的任务(如回忆过去发生的事件)变得容易一些,我们常常设法简化一个比较复杂的现象或判断(如给不同的满意程度打分),或将一个调查题项剪裁成宜于研究的问题(如只关注在某一特定时期发生的事情)。下面是一张调查问题最常见的组成部分和它们的一些实例的清单(本清单并未穷尽调查问题的所有组成部分)。

涉及的具体时间

在很多时候,研究涉及的事件只发生在某一特定的时间框架内。例如:

这学期或者说今年一月以来,您经常去图书销售中心吗?

或

您认为在去年,您的人身或财产是否受到过犯罪行为或犯罪企图的侵害?

在大多数情况下,我们确定的时间必须限于被调查人有可能提供精确的行为计数或事件计数的范围之内。确定什么样的时间框架适合于大多数被调

查人,取决于我们所了解的其他调查所采用的框架和我们对自己的调查与这些调查的相似程度所做出的判断。

在另一些情况下,出于某些与研究有关的实际考虑,我们所感兴趣的时间只是某一特定的时段。如果我们希望将去年人们有关犯罪问题的态度与来自以前的研究的数据进行比较,那么我们就必须将被调查对象报告的时间限制在去年。

概括性判断

有时,为了避免做过多的限定,或考虑到问题的复杂性,我们要求被调查人做一个概括性的判断。

一般讲,您认为在您居住的地区,犯罪行为是很严重、比较严重、不太严重或完全不严重?

总的说来,你对你的指导教师是否满意?

没有这些限定,有些被调查人也许会不知道怎么来回答提出的问题,譬如提供诸如有些类型的犯罪有所减缓而另一些类型的犯罪却更加严重这样一类答案。无独有偶,有些学生也会提供诸如有时他们对自己的指导老师感到很满意,但有时却不怎么满意这样的答案。这些限定的目的在于明确一种"全面地考虑过的"判断类型。

形容词和其他限定词

正像我们可以将答案限制在某一特定的时间之内一样,我们也可以将它限制在某一特定的地方(确切地或一般地加以定义)或多种事件中的某一种类型。

您认为在您居住的地区,去年暴力犯罪方面的情况有所好转还是更加严重,或没有什么变化?

当研究的目的在于对同一问题的各个不同方面,如对不同地方的不同犯罪类型进行比较的时候,这种方法是经常被采用的。

行为的原因

如果研究的问题或假设旨在探究行为的原因(或比这个还要困难得多,旨在探究被调查人的观点),那么问题必须力求使这些原因变得明确。

出于对犯罪问题的担心,为了自卫,您有多少时间随身携带武器?

作为一个练习,请将问题"您瞧过多少次医生?"转写成四种不同版本:①增加一个具体的参照期;②只问及为同一情况去看医生的情况;③只问及求诊某一科医生的情况;④包括所有前三种情况。为了使练习更有意义,请不要

套用前面已经给过的例子。

答案类别

一般情况下,答案的类别都是一目了然的。答案类别的选择,在很大程度上决定了我们后面看到的数据的形式。在讨论影响答案类别的某些因素之前,我们先来看一下答案所能测度的对象的种类。

调查问题可以测量的对象是很多的:态度问题,如同意还是反对流产、对当地警察的办事效率和对工作的满意程度等;行为问题,如看电影的频度;个体属性或事实问题,如年龄、收入或种族;知识问题,如后天免疫缺失综合症(艾滋病)的传染途径等。调查访谈得到的诸如这样的量度的质量好坏,不仅取决于问题表述的精确程度,而且也取决于答案的类别是否合适。有时恰当的答案类别是问题的自然而然的产物。例如,在问"在刚刚进入马里兰大学之前,您在什么样的学校就读?"这一问题时,我们只需考虑学生可能就读过的地方就可以了,如:

高中
从一所两年制的大专转来
从一所四年制的大专转来
休学之后返校

如果我们无法确定给出的答案类别中是否已经包含了所有可能的类别,我们可以增加一类"其他地方(请具体说明)"。在一个问题与一种事实或属性有关时,在大多数情况下,可供选择的答案类别必定都是很清楚的。而在问题与观点有关时,答案的类别,甚至维度可能都是不十分清楚的。例如:

有些学生认为,根据案例具体情况的不同,对拥有和/或吸毒的学生的处罚,应该是给他们一次进行两年的随机毒品检查的机会,而不是直接取消他们的学籍。你对这样的政策是很满意、比较满意、不太满意,还是很不满意?

我们的关注点可能都集中在我们所使用的"满意"的分类上,但是我们也有可能在某些更为基本的问题上犯了错误。"满意"是我们想要的词吗?那些对这一政策有一定经验的学生,可能说的是他们对这一政策实施的结果的满意程度,但是这与我们问题的本意有一定的出入;而其他没有这方面经验的学生,即使能对满意不满意给出答案,也无精确性可言。而也许我们真正想要知道的问题是学生认为这一政策是不是公正的或合适的。

在某些情况下,答案的类别就在问题之中,我们只需列出对选项的编码。一个简单的例子是"您在课程结束之后,卖掉还是保留您用过的书?"显然这一问题的答案只有两类:"卖掉"和"保留"。但是即使在这样比较简单明了的情况中,我们也必须确定答案是否已经包括了所有的可能性。如果我们更加

仔细地考察一下这个问题,就会发现,学生也可能选择“卖掉一部分和保留一部分”,这样我们就需要增加这一选项。是否还有其他什么可能性呢?

一个更为复杂的答案与问题一体的问题是有关犯罪研究的。假如我们希望了解被调查人对罪行判决的不同选项的想法,即了解对除了送进监狱这一判决之外其他判决的想法,那么我们可以设计编写诸如这样的问题:

您认为犯有抢劫罪的人,应该被送进监狱,还是需要每月向假释官报告一次、每天向假释官报告一次,或受到电子监控,以便随时都可以知道他在什么地方?

此外,我们应该懂得,答案的类别是问题的核心,因而我们必须以此为中心来设计编写问题。但是在许多情况下,在问及态度、等级排序或对问题的评价、行为等问题时,我们可以有很多种方法来构建答案选项。表 4.2 列出了一些常用的、测量态度、行为和知识的答案分类。虽然这些分类远远不是这些方面答案的分类的全部,但它却可以作为答案分类的基础(量表和问题构建的更多例子,请参见:Sudman & Bradburn,1982)。大家注意,这些分类中的某一些,很容易应用于其他的维度。例如“非常/比较/不太/完全不”等量词固然可宜于很多量度,但尤其宜于诸如“有帮助的”、“严重的”、“有关的”、“可能的”和“有趣的”这样的量度。它们彼此的排列是有序的,但它们的测量等级的选择则取决于被调查人的判断;某个人的“好”可能是另一个人的“一般”(关于模糊的规定请同时参见:Bradburn & Miles,1979)。

表 4.2 列出的答案分类,大多数都为三或四类,不过这样的分类数并非自然的分类。我们建议采用较少的分类,尤其是在电话调查中,更加应该如此。因为被调查人在确定自己的答案时,必须要记住答案的分类。同样,如果对调查的主题没有强烈的感受,许多被调查人可能选取容易的答案,即中性或“无所谓(no opinion)”类的答案,然而,这些被调查人中的许多人会倾向于取趋向于某一或另一方向的答案,从而使最终得到的数据比较有用处。正因为如此,在试调查之后,如果我们发现被调查人都倾向于取中间的选项,那么明智的做法是在修订问卷时考虑提供一个中间类的答案。有关分类数和中性类别的讨论可谓汗牛充栋,不一而足,但是它们都超出了本书讨论的范围(限于篇幅,有关这方面的详细的讨论,我们不准备在这里对它做更多的讨论,有兴趣的读者可参见有关的专注和文献(有关这方面的详细讨论请参见 Sudman & Bradburn,1982;Schuman & Presser,1981)。

表4.2 常用答案分类量词

观 点

完全满意/基本满意/还算满意/不满意/很不满意
很重要/比较重要/不太重要/一点也不重要
同意/反对
强烈反对/反对/支持/强烈支持

知 识

很熟悉/比较熟悉/不太熟悉/一点也不熟悉
真/假
很多/有些/很少/一点也没有

事件或行为的频数

从不/每学期不到一次/每学期一次/每学期两次/每学期三次/每学期三次以上
每天/每周/每月/每年/从不
总是/经常/几乎不/从不
总是/有时/从不
所有/大多数/有些/少数/没有
经常/有时/很少/从不

程度排序

变好/变坏/几乎没有变化
优秀/好/一般/差
大大高于平均/略高于平均/平均/略低于平均/大大低于平均
很公正/公正/不公正/很不公正
高/中/低
小/中/大

还有另外一种形式的量表,量表上面只标出了两个端点的值,下面就是一个这样的实例。

在一个含1到10,10个值的量度校园犯罪严重程度的量表中,1表示一点也不严重,而10则表示极度严重。则校园内犯罪有多严重?

这一方法可以得到比较有区分度的答案(有变化的答案)。然而,它也的确使分析变得复杂了。为了表明犯罪的确已经成为了一个问题,我们是否需要计算平均得分?将量表的不同部分合并成高、中、低三类?或使用某种最低限度(如任何大于7的得分)?所有这些处理方法和其他的处理方法都是可能的。如果量表中存在许多测量点,测量点又是一目了然的,且工具是自填式的,这样被调查人使用起来就会比较容易。

我们应当对观点性问题的主观性和固有的复杂性有所认识(有关如何处理问题的详细介绍,请参见 Turner & Martin,1984)。在两个被调查的学生独立回答"您对图书中心的预定服务满意吗?"这一问题时,他们对满意这一词的理解可能有一定的差别。这就是说,满意这一维度含有多个方面,而不同的人(或在不同的情况下)对不同方面的关注有所不同。其中一个学生,他的满意程度可能主要来自某次与一个出纳的不愉快的遭遇,而另一个学生,他的满

意程度可能主要取决于货物的平均价格。固然,每一个被调查人都理解满意这一词,但理解的角度却各不相同。

从这一观察中,我们还可以得到其他一些有关问卷设计和访谈的经验教训。首先,如果被调查人能正确地理解满意度这一点是研究的关键,如本例就是这样,那么我们必须采用多重问题。在得到了普遍的量度值之后,我们紧接着还应该问一些题项。在这一例子中,我们可以紧接着问一些这样的问题:"您(满意/不满意)的主要原因是什么?"

其次,被调查人也许想知道我们希望了解的是哪一方面的满意程度。在他们问调查员这一问题时,调查员的标准答案应该是:"无论您认为的是哪一方面都可以。"这就是说,我们应该让被调查人根据自己的认识进行选择,无论他们选择的是什么(Fowler & Mangione,1990)[3]。在诸如这样的场合,尤其是在被调查人对调查的主题不甚了解的场合,调查员很可能在不经意间影响了被调查人的答案。而这也正是我们之所以强调调查员必须照本宣科地读问题和禁止使用解释性的"有帮助"的评论的又一个原因。

最后,我们在另一个似乎比较简单的情形中,再一次看到了潜在的复杂性,因而我们可以考虑将探讨被调查人是如何在特定的问题情景下理解诸如"满意"、"严重的"和其他词汇的方法加到我们的试调查计划中去。这也许也会对试调查方法的选择有所影响。我们将会在第 6 章再次对这一问题进行详细的讨论。

在构建事件或行为发生的频数类别时,我们必须同时考虑类别个数和发生次数。而究竟如何确定则取决于我们对某个时段内可能发生的事件或行为的多寡的判断。如果我们高估或低估了某一频数,那么大多数被调查人就会被归并为一类,而我们就得不到希望得到的信息。在这样的情况发生时,我们就说量度的差异很小或没有差异。表 4.2 中的以学期为测量单元的答案量表(在条目"事件或行为的频数"下),虽然对一个有关与指导教师见面或逛书店这样的问题可能效果不错,但对去图书馆的次数的有关问题则可能效果不佳。因为大多数学生每学期去图书馆大大多于三次,这样,许多被调查人将可能被归并成一类。答案量表的选择还有另一种情形,在这种情形中,主导问卷编制的是我们对世界(在目前的情形中,世界便是学术界)是如何运作的认识。通过试调查,我们对我们自己的认识进行评价。

有时,我们甚至对如何才能提供恰当的选择这一问题毫无头绪。这时,我们可以在试调查中选用开放式问题(没有确定的答案选项),以了解行为发生的可能频数(或其他类别)。然后我们再根据这些信息构建用于正式调查问卷的封闭式问题。例如,如果我们不了解在一般情况下学生与指导教师见面的次数是多少的话,那么试调查的问题只要像"上学期你总共见了你的指导教师多少次?"这样就可以了。根据试调查中了解的次数,我们就可能建立诸

如"从不"、"一学期一次"、"一学期两次"、"一学期三次"或"一学期三次以上"这样的答案分类。另一方面,如果我们在试调查中发现,许多学生一学期见指导教师的次数在三次以上,那么在构建一组分类时,我们就可以将这一情况考虑进去。

在试调查之后,如果我们还是无法确定频数的范围究竟有多大,那么最好还是使用开放式问题,在调查完成之后再对答案进行编码,即使在正式调查中也应该如此[4]。例如在遇有一个问题只问及三种被调查人,而实际的被调查人的类型远不止这三种时,这样的情形就有可能出现。因为在这样的情形中进行的试调查所收集到的信息,一般都无法满足构建答案类别的需要。

确定调查问题的不足之处

与普通的作文一样,编写调查问题在很大程度上是一个不断修改的过程。为了在随后的版本中能使问题有所改进,我们必须首先用批评的眼光审查每一个问题,找出它们的毛病。多方面求教,无论是向一个人还是一群人求教,都对问题的编写有很大的帮助。我们编写的大多数问题可能不是为专业人士所用,因此文字必须一目了然,并能为一般的读者所理解。我们经常会遇到这样的情况,五个人都仔细地阅读了我们的问卷,其中多数人都会发现其中存在的问题,但每一个人发现的不足或问题,可能是其他四个人所没有发现的。几乎所有的问卷编写工作都会从多个读者的阅读中受益。

当然,我们作为设计者,可能并不认为"不足"是不足,但是我们因此而有了与我们的"读者"讨论问题的基础。一旦我们发现了问题中存在某个不足,我们就必须采取措施加以改进。这样我们就能对问题进行修改,在保留测量目标的同时,去掉不足之处。在最终达到我们的满意之前,一个问题被修改了多次是常有的事。

对问题进行有效地修改是很困难的,但是一旦我们真正发现了一个问题,那么经修改后的题项,即使不能说是尽善尽美的,至少也可以能得到一定的改进。即使我们并不完全满意修改后的版本,至少我们也有了一个良好的开端。当然,无论我们的工作做得多么精细,也难免留下某些不足之处。正因为如此,试调查才是那么的重要,即使对那些最有经验的调查研究者来讲,情况也同样如此。

某些问题结构的麻烦是问题本身所固有的。我们建议避免采用的问题形式之一是同意—不同意型问题。在这种类型的问题中,我们在给被调查人一个陈述的同时,询问他们是否同意这一陈述。通常我们提供的答案的种类为二、四或五类。下面便是一个同意—不同意型问题的例子:

我将要问您几个有关刑事犯罪司法系统方面的问题。请您告诉我您对每一个问题的态度是完全同意、同意、不同意,还是完全不同意。

警察的工作做得非常出色。您对这一说法是完全同意、同意、不同意还是完全不同意?

虽然这样形式的问题使用非常普遍,但是它的确存在不少问题,并一直受到调查方法论专家的强烈批评(Converse & Presser, 1986)。研究显示,不论问题的内容如何,被调查人都有一种同意的倾向[叫做"默许倾向(acquiescence response set)"]。不仅如此,我们还可以证明(Schuman & Presser,1981)这一倾向与文化程度有关,文化程度较低的被调查人较之文化程度较高的被调查人更易于倾向同意。

另一种必须避免的问题形式是双管式问题(double-barreled question),采用这样一种形式的问题,往往是不得已而为之。这种形式的问题有两个部分,在被调查人看来每一部分可能有所不同。例如,"您认为警察和法庭的工作做得很好、好、一般还是很差?"如果一个被调查人认为警察的工作做得很差,但法庭的工作却做得好,他就无法回答这一问题。与同意—不同意型题项不同,双管式问题并没有反映某种偏倚或研究者的偏好,相反它的缺陷是结构上的。对于同意—不同意型题项的不足,专家们的观点不尽相同,但对于双管式问题的看法,专家们的意见并无分歧,他们都认为这种形式的问题是不可取的。

解决的方法无非是将原来的题项拆分成两个问题。双管式问题常常是因想节约问卷的篇幅而将多个专题捆绑在一起而引起的。其结果是使问题无法理解和回答,进而使得到的数据无法用于分析。

最后,也是调查问题中最难以对付的问题是含混不清。在被调查人身上测试某个问题时,我们发现他们对问题理解的差异是如此之大,以至于看起来这是有意而为的。如果我们从一开始就去除了问题的模糊性,那么后面的所有工作就会变得比较容易。佩恩(Payne,1951)在他的经典著作《提问的艺术》(*The Art of Asking Questions*)一书的结尾,为我们提供了一张在编写问题时应当考虑的一些问题的、行之有效的检查清单,其中许多内容都与模糊性有关。

注 释

1 该电话调查由马里兰大学调查研究中心(the University of Maryland Survey Research Center)主持。它由几个有关学生的调查的独立研究组成,调查由马里兰大学的学生实施。

2 一般讲,"不知道",或简称"dk"这一答案的编码为 8(或 88),而拒绝回答的编码为 9(或 99)。这种方法能使整个问卷的这一类结果都使用相同的编码,不论答案的类别数是多少,这样将使分析工作简化。在这里列举的例子中,并没有提供"不知道"这一选项,但被调查人也许希望选择这一选项。

3 注意,这一策略与我们在调查有关属性或某种事件或行为发生的事实性问题时所采取的相反。在那样一些问题中,调查的"目标"已经做了清楚的定义,如轿车的所有权,而我们希望确保的是,不同的被调查人能在同一意义上报告自己的情况。

4 注意,这种形式的开放性问题将得出一个单独的数字,或某一时期内的次数,因此不存在那种开放性叙述中存在的编码困难问题。

问卷设计:编排问卷

第5章

Questionnaire Design:Organizing the Questions

一般讲,问卷都被编排成若干部分。编排的逻辑依据是抽样设计、数据收集的方法步骤和问题的使用方式。大多数问卷都包括调查简介、被调查人的选择方法、实质性问题、背景或人口学问题等内容。有些问卷在这些内容之外,可能还包括几个为数不多的后调查题项(post interview items)。尽管这些部分的具体形式和编排方法将因特定的调查和使用方式而异,但是一般讲,它们的基本目的将如下面所介绍的那样并没有什么差异。

调查简介。在我们因调查而与被调查人(或住户中的知情人)联系的时候,我们必须提供有关调查的足够信息,以取得他们的配合。这一个步骤是不可或缺的,即使我们事先已经就我们的调查发过通知。

被调查人的选择。这一步骤通常是调查总体为一般人口,且分析单位又为个体的调查所不可或缺的。一般讲,我们采用一阶或多阶抽样方法,从一个抽样框中抽出了住户单位,然后,我们再在抽出的住户单位中随机地抽取被调查人。在一个分析单位为个人的研究中,如果我们未对被调查人选择的方法做过周密的考虑,那么便有可能导致样本产生严重的偏倚。例如,户中最可能接电话或开门的人,较之户内其他合乎调查要求的成员更有可能被选作被调查人。

实质问题。在问卷的这一部分,我们提出的问题涉及了我们研究目标中的每一方面。它是我们问卷的核心,是我们收集的数据的重点,因此我们必须把主要的人力和物力放在这一部分。

背景问题。在一般人口总体的调查中,我们常常需要收集一些背景信息,这方面的信息一般都是有关被调查人的人口学方面的信息。虽然诸如这样的问题并不存在一组标准的问题,但是性别、年龄、种族、教育、婚姻状况和收入都是经常被问及的问题(有关若干常用的人口学题项的格式,参见附件B中的犯罪调查问卷)。我们之所以要收集这些数据,原因不外乎三个:第一,我们的分析可能需要它们。我们已经有了某些假设,认为它们将有助于我们对实质性问题存在的差异做出解释。例如,如果我们认为在白人和有色人种之

间,在对待刑事犯罪司法系统的效力问题的态度上可能有所不同,那么我们便需要收集有关每一个被调查人种族方面的信息。其次,我们可能需要将我们的调查的人口学特征的分布与普查数据做比较,以确认我们的样本的代表性。第三,如果我们决定采用后分层加权(poststratification weights)(参见第 7、8 和 10 章中有关抽样和报告撰写的讨论),那么这些数据同样也是必不可少的。

对于特定总体,如学生的调查,根据研究的需要,问卷中可能也需要包括某些背景问题。例如,如本书附录 A 中刊出的学生综合调查的问卷,可以问一些有关班级名次、年级平均成绩和其他诸如这样的问题。但是在加入诸如这样的题项之前,我们必须清楚,我们将如何使用它们,因为增加背景问题,势必会增加问卷的长度,从而导致调查费用的增加。

后调查问题。在调查全部完成之后,我们可以问被调查人和调查员,或同时问这二者几个额外的问题。这样的问题一般有两种类型:第一,我们希望收集有关刚刚完成的调查的信息。例如,我们可以要求调查员对被调查人对某些问题的理解程度或对回答某些题项的自愿程度做出评价。同样我们也可以要求被调查人告诉我们,他们是否对某些问题不甚理解,或告诉我们他们为什么拒绝,或无法回答某些问题。诸如这样的后调查问题,不仅在对某些可能比较困难的、过于敏感的工具进行试调查时特别有用(参见下面有关试调查的讨论),而且也可用于我们的正式调查。然而,我们必须记住,后调查题项会增加我们的调查费用,因此不可盲目地把它们加入到我们的问卷中去。它们也必须像其他问题那样,通过有用性检验。

第二,有时在调查完成之后,如果我们准备再次与被调查人联系,或将调查结果送给被调查人,我们可能要求他们告诉我们第二信息源的姓名(在被调查人住址变动时,知道他们的下落的亲戚或其他知情人)。

过渡性和辅助性信息。除了这些问卷的主要部分之外,有时我们需要在被调查人回答我们的问题之前,给他们提供某些信息。例如,介绍一下我们要求他们选择的两个社会服务项目。在另一些场合,我们需要编写一些过渡性陈述,以提前对一个新的专题做一些简单的介绍。

提供辅助性或过渡性信息的做法,有可能导致某种偏倚。在给被调查人提供有关信息以帮助他们回答问题时,我们必须加倍小心,以免被调查人的答案产生偏倚。对于一般公众的调查,常见的偏倚似乎是有意而为地引导对被调查人答案的影响。当然,这样的偏倚的确存在,但是我们在无意之中导致的偏倚却更为普遍。过渡性陈述编写不当,同样也会导致偏倚。这样的陈述应该简短、价值中立,它是沟通问卷中两个部分的桥梁,提醒被调查人,他们在下面将要回答的那些问题与一个新的专题有关。

过渡性陈述的另一个作用是在进入某一含有敏感性或令被调查人感到不安的题项,如有关性行为或某些非法活动的题项部分的调查前,重申调查的保

密性原则。当然,我们必须在调查开始之前就做出保密的承诺,但是在调查的关键时刻,再次重申这一承诺,无疑会对降低题项的无回答率有所帮助。

在提供有关信息或问卷的过渡性陈述中必须注意的问题如下面所列。在每一个必须注意的问题之后,我们都辅以一个过渡性陈述或章节简介的实例。

- 社会愿望。"学生公共医疗服务项目是一个正在发展过程中的新项目,其目的在于使学生了解酗酒的危险性。以下是一些有关您的饮酒习惯的问题。"

有关饮酒的问题或多或少都有一定的敏感性。报告过量饮酒的行为与社会愿望相背。这一过渡性陈述非常有可能增加问题的敏感性。被调查人有可能因此低报自己饮酒的次数和数量。

- 带倾向性的形容词。"去年,为了改进学校指导老师提供的咨询服务,政府任命的指导老师一直在努力工作。这种咨询服务是对学生指导工作的一个重要组成部分。本调查的目的之一是了解您对他们在咨询服务方面的工作表现的看法。"

过渡性陈述和帮助信息同样可能使被调查人在随后的问题中给出正面的评价。当然并非所有的学生都如此,但是有些学生,特别是那些事先对咨询服务没有明确看法的学生可能会因此而受到影响,给出正面的评价。

- 与权威有涉。"大学校长和理事会提出了一个重组学术咨询系统的计划。某些团体对这一计划持反对意见。下面的问题将问及您本人对这一计划的意见。"

在这一例子中,一方面是声誉显赫和位高权重的管理者,而另一方面则是人微言轻的反对者团体。诸如这样的陈述势必导致被调查人在评价这一计划时举棋不定。

为了避免过渡性陈述和背景信息可能导致的偏倚,我们可以采用一些什么样的方法呢?最简单的办法是尽可能地减少它们的使用。在很多情况下,它们实际上并不需要使用。其次,避免使用感情色彩强烈的词汇,调查不需要任何感情色彩。不要对任何问题表态。在的确需要使用过渡性陈述和帮助性信息时,务求简明、中庸和平和。最后,在问卷测试阶段,我们不要忘记对过渡性陈述和帮助性信息的测试,须知它们也是问卷的一部分,因而也应该以同样严格的要求进行评估。问卷中的每一部分都有可能对答案产生影响,我们必须像对待其他问题那样对待它们。

调查简介

调查简介有多种作用。它是对调查的扼要介绍,告诉被调查人调查的内容、目的、组织者和其他若干有关调查的详细信息。它给预期的被调查人提供了有关调查的充分的信息,以促使他们在知情之后接受我们的调查。它最重要的作用是吸引预期的被调查人参与我们的调查。

而所有这些目标必须以最快的速度实现,特别是在电话调查中。因为在电话调查中被调查人本来就易于拒绝不请自来的电话提出的要求。不仅如此,更为重要的是,我们还必须认识到简介中的每一个字和每一段话可能都会对三个目标有所影响,因而我们可用这一点来达到我们的目的。例如一个简单的介绍性语句:"我们正在为马里兰州的首脑部门做一个有关街头暴力犯罪的调查。"在传达了调查目的、组织者这些信息的同时,还告诉了被调查人调查数据将为州有关部门使用,可能作为处理与犯罪有关的一些问题的参考。

问卷设计的一个目的是使可能的被调查人懂得调查对他们是很重要的,他们为之付出的时间和精力是值得的。我们希望被调查人能严肃地对待我们的调查,并提供完整和精确的答案。确保被调查人能做到这一点,这是我们的第一个任务,而在我们介绍我们的调查的时候,这一任务便已开始付诸实施。

在过去的若干年间,参与问题已经变得尤其重要。一般人口总体调查,特别是电话调查的回答率严重下降。尽管人们对下降的原因见仁见智,观点不尽相同,但对下降已经发生这一事实却没有什么异议。在设计和实施一项调查的时候,我们必须在考虑调查的正式内容的同时,周密地考虑如何使被调查人参与我们的调查。如果回答率低到有损于我们的调查的可信性(credibility)的程度,那么不论我们如何精心地设计我们的工具的其他部分,其价值都不会太大。

可信性与实际的无回答偏倚或低调查信度(reliability)不同。在回答率趋于下降的时候,我们则担心无回答偏倚会有所上升。然而,我们通常无法对这样的偏倚进行测量,因为这仅仅是一种可能的(potential)效应而已,它也许存在,也许不存在。某些研究表明,在某些特殊的情形下,低回答率并不一定带来严重的无回答偏倚(参见 Keeter, Miller, Kohut, Groves, & Presser, 2000; Curtin, Presser, & Singer, 2000)。不过在与之相反的情形下,即在通常的情形下,我们则没有证据证明低回答率不会导致严重的无回答偏倚。换言之,在通常的情形下,低回答率将会降低调查结果的可信性。

除了其他问题之外,保持被调查人的参与始终是我们在调查过程中不可忽视的问题,而目前,在起草调查简介的时候,我们也必须对这一问题给予很

多的关注。与之同等重要的问题是调查员的工作能力。调查员是否能在介绍调查的时候,回答被调查人提出的问题和解除他们可能存在的顾虑对保证调查的质量也是非常重要的。当你在阅读下面有关如何做调查简介的忠告时,不妨试着设想简介的每一部分可能在一个可能的被调查人心中引起的问题,并进而设想我们应该如何引导调查员回答每一个问题。

重要的问题在于我们必须认识到调查的主题本身(survey topic per se)和我们的特定研究(particular study)二者之间是有区别的。只是将我们调查的主题告诉被调查人是不够的,我们似乎还必须用一些时间把我们的研究给被调查人做一些介绍。例如,一个被调查人可能对犯罪问题非常感兴趣,并认为这是一个非常重要的社会问题。然而,这并不一定意味着被调查人也认为我们所做的有关犯罪的研究就是很重要的。

限于篇幅,我们不可能在问卷中对"可能的被调查人"反复强调这一点,但是我们必须在心中牢牢地记住这一点。每一个被选入样本的人既是一个可能的被调查人,但也可能是一个可能的拒绝接受调查或其他类型的无法进行调查的人。

调查简介可以用一封随问卷邮寄的封面信件、打电话之前的预达信件或在个别访谈时送给被调查人;或当调查员在被调查调人家中与他进行"冷"接触时,进行调查简介时面交。我们将在这里依次介绍有关这些初次接触形式中的每一种中比较重要的方面。首先,我们先来介绍一些有关导向性的问题。

在考虑如何起草适合于每一种调查方式的简介时,我们应该弄清楚被调查人经常会对将要进行的调查考虑一些什么问题。这些问题包括:

- 调查的内容和目的是什么?
- 做调查的是什么人?
- 调查的组织者是谁?
- 调查为什么这么重要?
- 调查的结果会用来做什么?

这些问题是一个被调查人在同意接受调查之前经常问到的问题。被调查人一般都希望对这些问题有所了解似乎是不言而喻的,但是许多研究者却对这样一个对被调查人是否接受调查起决定性作用的常识性问题视而不见。被调查人关心的其他问题则包括:

- 调查与我有什么关系?
- 调查为什么选中了我?

有些被调查人可能还想了解有关调查的更多的情况,但是限于篇幅和时间我们可以传递给他们的信息总是有限的。上面列出的那些问题是大多数一般人口总体调查中都会遇到的问题。我们不难理解,对于一个特定的调查,除

上述那些信息之外,我们还必须提供其他一些基本信息。例如,如果某一个调查需要问及一些敏感或有关犯罪行为的问题,那么我们可能必须对保密性做出特别的承诺。如果样本选自一张特殊的清单,如一个医疗卫生组织(health maintenance organization,HMO)的患者名单,那么明智的做法是告诉被调查人,我们是怎么知道他们名字的。

电话简介

我们怎么确定电话调查的简介应该包含一些什么内容呢?我们又应该怎么确定什么是将我们的简介传达给被调查人的最好方法呢?下面就是一份有关马里兰州的犯罪问题的随机拨号(RDD)电话调查的调查简介草稿:

> 您好,我叫×××,我从马里兰大学调查研究中心给您打电话。调查研究中心受州长 1991 号委托,进行一项有关街头暴力犯罪的调查。调查结果将用于规划建立预防减少犯罪的项目、青少年教育项目和进一步完善法律体系。您所在的户由随机选择方法选中。您是否接受我们的调查完全出于您的自愿,但它对我们的调查最终是否有代表性非常重要。您为我们提供的所有答案都将严格保密。

让我们先花一点时间来检查一下,这一份简介草稿是否已经包含了上面列出的那些可能的被调查人通常想要了解的所有问题。然后,在进一步阅读之前,我们再费一点时间来设想一下,如果我们自己在电话中听到这样一个简介,会有什么样的反应。记住,被调查人并没有想到会有这么一个电话,且我们事先也没有给他们发过信。他们对调查所了解的一切都来自这一简介。

所有重要的问题似乎都已经做了解释,但是却太长了。被调查人拿起了电话,并被要求聆听一段话。记住,我们希望把这些要点传达给被调查人,但是必须在很短的时间内。下面是我们的第一个修改稿。我们不妨将它与原稿做一个比较,并分析一下修改的原因。

> 您好,我从马里兰大学给您打电话。我叫×××。我们正在为马里兰州的首脑做一项有关街头暴力犯罪的研究。研究结果将用于帮助制定减少犯罪的计划。您所在的户经随机选择法选出。您是否参加我们的调查完全是自愿的。您给出的所有答案将严格保密。由于调查的需要,我们必须在您家中年龄在 18 岁或以上、接下来就要过生日的成员交谈。您家中哪一位满足这些条件?

首先,我们在自报姓名之前,先告诉被调查人我们从哪里给他打电话。记住,许多拒绝接受调查的事情,都发生在接触刚刚开始的时候。如果被调查人首先听到的是一个陌生人的名字,那么事情就大为不妙。但是如果首先听到的是一个组织的名字,且听起来是一个合法的组织的名字,那么接电话的人便

可能会认为打来的电话有什么要事相告,而不是打错了电话或电话推销。这样的事情固然不是什么大事,但是在问卷设计中,许多事情本身并不大,如果累积起来就可能成为大问题。

其次,我们用研究(study)一词代替了调查(survey)。有些证据表明(Dillman, 2000)调查一词含有某种贬义。此外,推销电话也常常伪装成调查。而许多被调查人都对这样的电话非常敏感。随着新的全美不接受电话推销名单(National Do Not Call List)的出现,这一问题成为了一个更为敏感的问题。首先,尽管法律对调查和电话推销有所区别,但是许多被调查人并不了解这一事实。像这样不了解两者在法律上存在区别的被调查人,也许会认为我们的调查违反了新的法律,从而使我们更难得到被调查人的合作。其次,某些缺乏道德的电话推销员已经将他们的推销活动装扮成调查,而这样的现象可能还会有所增长。甚至还存在这样的可能,某些电话推销员,可以装模作样地问几个他们并不打算使用的调查问题,作为一种在技术上遵守法律的手段。

关于研究结果使用问题的介绍我们则做了删节。第一稿在这方面的文字过长。我们在这一稿中也不再提及州长。提及州长固然能有助于得到许多被调查人的合作,但同时会减少那些反感州长的被调查人与我们进行合作的可能性。将调查组织资助者的地位限于马里兰州,可能会在最大程度上吁请人们的合作。

此外,有关被调查人的户是如何选择的和保密性的承诺对这一特定的调查本身而言可能并不重要。尽管我们没有说调查需要多少时间,但是如果被调查人问到了这一问题,调查员就必须说实话。调查员应该说明调查平均需要多少时间,但因为某些原因,实际上可能长一些,也可能短一些。同样,其他我们决定不写进简介中的信息,调查员也应该掌握,以在被调查人问及时随时作答。

我们将户内被调查人随机选择法写进了修订稿。它使我们从研究问题的介绍自然而然地过渡到向被调查人提出我们的第一个要求。通过这一要求,我们巧妙地结束了研究的介绍,在不知不觉间过渡到了调查的提问阶段。

下列最后一稿的简介似乎更加简短:

您好,我从马里兰大学给您打电话。我叫×××。我们正在为马里兰的首脑做一项有关街头犯罪的研究。我需要和您家中年龄在18岁或18岁以上,且接下来就要过生日的人说话。您家中有符合这样条件的人吗?

预达信件和封面信件

预达或封面信件是调查的重要组成部分。在调查员与被调查人联系之前,有时我们需要寄一封预达函给他们。而在互联网调查中,我们也可用电子邮件来发送预达函。封面信件通常与邮寄的问卷同时使用。虽然它们的关注

点有所不同,但是预达信件和封面信件的目的则大同小异。正因为如此,我们把它们放在一起讨论(有关这方面的完整的讨论,请参见 Dillman,2000)。

采用这两种信件的目的都是争取得到被调查人的合作。出于这样的目的,尽管在某些方面预达信件和封面信件的内容与调查简介十分相似,但在某些方面却有重要的区别。首先,与调查员填写的调查简介不同,预达或封面信件必须自达。第二,与调查员相比,信件更容易为人所忽视。正因为这些原因,这样的信件必须十分抢眼(但必须有专业水准)、达意清楚(但必须简短)和有说服力(但必须保持价值中立)。这样的信件必须能从大多数人收到的宣传品中脱颖而出,能使被调查人摆脱与我们进行合作的主要障碍,将研究者的意图传达给被调查人。

诚如迪尔曼(Dillman,2000)所言,这些信件一般以一种特定的顺序呈现以下内容:

- 调查的主题是什么,它的重要意义是什么,研究结果如何使用。
- 被调查人为什么对研究很重要。
- 被调查人是如何选取的。
- 承诺对信息保密。
- 在被调查人有问题时联系的电话号码和联系人姓名。

封面信件的长度不能超过一页,且应刊印在有个人签名的、有抬头的信笺上。

在邮寄式调查中,封面信件的作用是介绍调查,所以我们不宜在工具本身及问卷中重复这些内容。问卷本身应当给被调查人提供如何回答问题的简明提示。而工具的每一部分都应标以不同的字体,使被调查人能清楚调查内容的变化。

问卷应该以什么样的问题开始

在引导被调查人通过了调查简介之后,我们已经可以开始问第一个问题了。我们应该遵循什么样的原则来选择这些问题呢?有两件事情我们应该记住:第一,被调查人决定是否参与我们的调查是分阶段的。第二,问卷的题项或不同部分之间可能存在着某种逻辑关系。

有关如何促使被调查人参与调查的研究文献并不是很多(希望对这一问题有更为详尽了解的读者可参见:Groves,1989;Groves,Cialdini,& Couper,1992;Groves,Dillman,Eltinge,& Little,2002)。但是我们确实有很多证据可以证明,在调查员填写的调查中,许多拒访发生在调查简介或提出最初的几个问题的阶段。考虑一下被调查人依据封面信件或预达信件和调查简介中提供的

信息,决定参与调查(或至少是聆听我们的谈话)的最初的动机是什么,可能对我们会有所帮助[1]。然后,被调查人听取(或读取)第一个问题。除了开始得到调查数据之外,第一个问题还应该有什么作用呢?它应当有下列的作用或特点:

与核心主题有关。我们不应该在告诉被调查人我们的研究与犯罪有关之后,接着再问的第一个问题是有关教育年限的,如某人的上学年限。尽管这个问题可能与分析密切相关(且将在后面问及),但是许多被调查人会因此而问自己(或调查员):"我的教育年限和犯罪问题有什么关系吗?"正因为如此,一般我们不应首先提出有关人口学特征的问题。

很容易回答。大多数调查涉及的各种问题,对于被调查人来讲,回答的时候难度不尽相同。明智的做法是,以大多数被调查人都很容易回答的问题来开始我们的提问。在某些被调查人可能因为发现问题有一定的难度而跃跃欲试的时候,许多其他的被调查人则正在为自己是否应该接受调查犹豫不决,也许他们会因此而终止调查。大家必须记住,我们的目的是得到大多数被调查人的合作。作为一个练习,大家在读完这一章之后,读一下附录中的问卷A和B,并找出和标明那些你们认为比较容易回答的问题和那些你们认为比较难以回答的问题。

感兴趣。有关调查研究的文献鲜有直接论及感兴趣问题这一主题的。虽然某些有关调查员——被调查人关系的调查对这一问题已经有所接触,但是它们几乎都没有把它作为一个独立的因素来考虑。然而,常识已经表明,在其他条件相等的情况下,任何人都比较愿意回答比较有兴趣的问题,而不是平淡乏味的问题。一般情况下,人们都更乐于表明自己的观点("您认为在您居住的小镇,犯罪问题在上升、下降,还是没有什么变化?"),而不是事实("您每天都看报吗?")。

适用于大多数被调查人,并能为他们所回答。假如调查以这样一种方式开始,即以一系列许多人因为问题不适用于他们而只能回答"不知道"的问题开始的调查,几乎可以肯定是不能引起被调查人的兴趣的。如"一般讲,您认为大学图书中心销售的软件价格合理吗?"这样的问题就属于这种类型的问题。因为许多被调查人对这一问题的答案必然是不知道,因为他们从来没有买过软件。

封闭式。在某些被调查人正在为有机会回答一个答案没有限制的开放式问题而欣喜不已的时候,许多其他的被调查人则都在为不知如何开始回答调查的问题而犯难。如果在调查开始之前,被调查人不曾对调查的问题有所考虑,情况尤其如此。

虽然这些指南对我们确定以什么样的问题开始我们的调查会有所帮助,但是它们毕竟不是万应灵丹,有时我们还是需要根据具体情况有所变通。重

要的问题在于我们必须知道在什么样的情况下应该有所变通和在遇到这样的情况时怎么进行变通。最常见的需要变通的情形可能发生在需要做一定的筛选的时候。在调查的目标总体与抽样框的总体不完全一致的时候，在调查一开始我们就必须确定该户或个人是否符合我们对目标总体的定义。例如：

您家中有几个年龄在 18 岁以下的孩子？

____填入孩子数并继续提问

（没有孩子）调查结束

在提"筛选"性问题时，有一点我们必须注意，那就是我们不可明确地告诉被调查人，什么样的目标户符合我们调查的要求。某些被调查人会因此而选择使调查无法继续进行的答案（如一个有孩子的男子，在回答上面这样的问题时，回答"没有孩子"）。一种可以避免这种情况出现的方法是使问题保持平衡，使被调查人无法了解什么样的答案会使调查终止（如"那些住在本户的人中，有多少个 18 岁或 18 岁以上的成年人？有多少个年龄在 18 岁以下的孩子？"）。

在很多情况下，筛选性问题都不会带来什么问题。例如，有时目标总体是由某一特定地理区域内的住户组成的，如一个县或校区，我们在做筛选的时候只要问一下居住地就可以了。在另一些时候，筛选问题可能比较敏感，例如在我们只对某些收入类型的住户感兴趣的时候。在为数不多的情况下，诸如这样的筛选性问题，可能会非常敏感，以至使它们是否可行（在发生严重瞒报的情况下）都成为问题，例如在同性恋调查中，情况就是如此。

苏德曼和布拉德伯恩（Sudman and Bradburn，1982）认为敏感性既取决于答案，也取决于问题。例如，一个非法使用毒品的问题对那些不使用毒品的被调查人来讲，就不存在什么敏感性问题。然而在另外一些场合，例如对大多数人来讲，有关性行为问题都是比较敏感的。一般讲，调查研究方面的新手最好不要使用敏感的题项，尤其不要用敏感的题项作为筛选性问题[2]。在我们必须比较早就提出一些有一定敏感性的问题时，以下几点是必须注意的：

- 只问绝对必要的、尽可能少的敏感问题，以确定被调查人是否有资格作为目标总体定义的被调查人。
- 给被调查人交代清楚，为什么必须问这样的问题。
- 如果费用允许，可考虑在筛选性题项前，插入一个或两个开放性的不敏感的"缓冲"问题，以打消被调查人的顾虑。

最后，如果问卷的某些部分只适用于具有某种特征的被调查人，那么上述有关第一个问题的建议则必须有所变通。例如，问卷的第一部分与抽烟的人有关，那么我们则必须首先问被调查人，他们是否抽烟，然后我们才能决定我们是否要问有关抽烟的那部分问题。

将问题编排成章节

现在我们已经对如何编写问卷开始部分的问题有所了解,然后,我们必须确定如何编排其余的实质性问题。这方面的研究为数也不是很多,所以我们只能再一次提出几个我们应该注意的要点。与问卷开始的问题一样,在下面的问题编排中我们同样要注意相关性、方便性、有趣性和适用性这样几个要点。因为大多数调查的终止都发生在调查刚开始不久的时候,而造成这样的终止的原因不仅应使我们在选择第一组问题时予以重视,也应该是我们在整个问卷编排中重要的考虑因素。

除了这些因素之外,我们还必须考虑整个问卷中问题的内在逻辑和问题之间及各个部分之间的衔接。无疑,如果某些问题有赖于前面的答案,那么这些题项的次序则必须以那样的逻辑为导向。能使被调查人感觉到问题的流向或自然的进程也是很有用处的。无论是调查员填写的问卷,还是自填式问卷,在问卷设计时,在这些方面花一点功夫都是值得的。诚如迪尔曼(Dillman, 2000)所言,就邮寄式调查而言,被调查人应该能感到问题的展开是有条不紊的,并觉得自己很快就能完成问卷的填写。我们不希望出现与某一特定主题有关的问题散布在问卷的各个部分这样的现象。同样,在一组问题使用的是同一个量表的时候,如果我们能把它们编排在一起,这样被调查人在回答的时候就会觉得比较方便。

现在我们就将这些要点应用于我们的犯罪调查。表 5.1 列出了调查的主题。我们如何在问卷中对它们进行编排呢?

表 5.1　马里兰犯罪问题调查的主要问题

1. 背景(人口学)问题
2. 个人犯罪经验
3. 对刑事犯罪司法系统的看法
4. 被调查人和他们的家庭对受到犯罪行为侵害的忧虑
5. 对州和邻里地区暴力犯罪问题的看法
6. 对州和邻里地区非暴力犯罪问题的看法
7. 使自己免受犯罪行为的侵害做过的一些事情
8. 对于备择的判决的看法
9. 对犯罪问题的一些思考,一般讲,范围限于邻里地区和本州

我们希望问卷中的第一组问题比较容易回答。无疑,对于被调查人来讲,比较容易回答的问题是他们以前考虑过的问题,而不是以前没有考虑过的问

题。有些被调查人可能曾经思考过有关刑事犯罪司法系统的问题,但许多其他的被调查人则可能没有。此外,正如我们已经了解的那样,这一组题项在回答时有一定的难度。有关犯罪的个人经验的题项应该满足便于回答这一条件。同样,有关被调查人专门为免于受到犯罪行为侵害做了一些什么这些题项,也必须满足这一条件。这两个问题或多或少都有一点敏感。作为一种选择,我们也可以适当考虑一下有关犯罪的问题。无疑,人们都会比较认真地考虑犯罪问题究竟有多么严重这一问题,且这个问题也与我们调查的主题密切相关。鉴于上述分析,我们认为,有关犯罪问题的思考可以作为我们调查开始的问题。

人们更有可能考虑到的犯罪问题发生的地方是自己居住的地区,而不是其他什么地方。因此我们的问题可以首先从邻里地区的犯罪问题开始,然后再问及州内的情况。那么我们是否应当首先问暴力犯罪问题,然后问其他类型的犯罪,最后再问一般的犯罪情况呢?或者,我们是否应该以别的次序来问不同类型的犯罪情况呢?我们根据什么来做出决定呢?有时一个问题的答案可以对一个实质性题项的答案有所影响。这就是说,答案会受问及的问题的上下文的影响。诸如这样的上下文效应一般是很难预料的,但是我们不仅必须想到它们是有可能发生的,而且还要判断它们有可能在什么地方发生。

例如,一个一般性判断可能会受到已经给出了详细评估的上下文的影响。在要求被调查人做一般性判断时,应使他们免于受自己已经给出的有关详细答案的影响。苏德曼和布拉德伯恩(Sudman & Bradburn ,1982)把这一现象称为"冗余效应(redundancy effect)",并指出了消除它的方法:"因而在一般情况下,如果我们要问一系列问题,其中有一个是一般性的,而其余都是特指的,那么最好先问那个一般性的问题"(p. 144)。然而,实际情况还要更加复杂,研究表明,在一个特指问题先于一个一般性问题时,许多被调查人有可能将有关这一问题的内容从他们给出的一般性问题的答案中排除出去(Strack,1992;Tourangeau,Rips,& Rasinski,2000)。在一组特指问题先于一个一般性问题时,一般讲,被调查人在回答一般性问题时,往往将各个特指问题的答案合在一起作为一般性问题的答案(Schwarz,1991)。

鉴于上述几点,我们也许可以先问有关犯罪的一般性问题(I),然后再问暴力犯罪和非暴力犯罪问题(E,F)。因为人们的观点会受个人经验的强烈影响,所以在调查的时候,对于被调查人来讲,回忆任何与犯罪有关的经验可能比较容易。我们不要忘记,在回答问题时,被调查人的记忆作用很大,因此在调查中,这里是一个问他们任何有关犯罪问题的经验的好地方(B)。少部分被调查人或他们的家庭(尽管多于我们猜想的)将会告诉我们,他们受到过犯罪行为的侵害,但是这一问题会引起他们对于受犯罪行为侵害的关注(他们可能会想起曾经在报上看到过的一个犯罪事件,或他们的朋友遇到的某些事

情)。这时,问及有关对于犯罪问题的担忧的问题就是一件再自然不过的事了(D)。然后,因为使自己或自己的家庭免受犯罪行为的侵害可以做的事情,必须基于或至少在一定程度上取决于自己对有关犯罪问题的思考、经验和期望,所以在接下来的部分似乎应该问有关这方面的问题,无疑,这也是顺理成章的(G)。

最后,我们剩下的问题是有关刑事犯罪的司法系统和备择的判决。我们以什么样的次序来问这两方面的问题呢?这时,我们再一次遇到了一般性问题和特指性问题。我们还是以苏德曼和布拉德伯恩(Sudman & Bradburn, 1982)的建议为根据,首先问有关刑事犯罪司法系统的一般性问题(C),然后再问判刑问题,即这一系统的特定部分(H)。正如我们将要看到的那样,回答备择判决问题也多少有一点困难,为使论证更有根据,把它们放在最后。与一般的电话调查一样,我们以人口学题项(A)作为我们问卷的结尾。最后的顺序(至少在进行试调查之前)如表 5.2 所示。

表 5.2 马里兰犯罪调查主题的最终排列顺序

对犯罪问题的看法,一般在邻里地区与州的范围内
对暴力犯罪问题的看法,在邻里地区与州的范围内
对非暴力犯罪问题的看法,在邻里地区与州的范围内
个人在犯罪问题方面的经验
对本人和自己的家庭受犯罪行为侵害的担忧
防止自己受犯罪行为侵害做过的事情
刑事犯罪司法系统的一些观点
对备择判决的一些观点
人口学问题

在讨论了问题的顺序问题之后,我们再转向为满足不同的调查组织需要的、调查题目范围很广的本科生调查。该调查并非始于一组研究的问题,这一研究的主要目的是收集一些学校不同管理部门制定计划所需要的信息。诸如这样的多目的研究叫做综合性调查。

问卷的设计满足以班为单位的样本做自填式调查的需要。答案是匿名的,而非保密的。与保密调查不同,在匿名调查中,研究人员无法通过姓名、电话或地址一一认定被调查人。匿名调查使任何人,甚至包括研究者都无法将答案与被调查人一一联系起来。在调查的问题相当敏感或涉及非法行为时,我们常常需要采用匿名调查。假定表 5.3 列出的主题决定了它们可能的排列顺序,那么本书后面的附录 A 中刊出的该调查的问卷,就可以看作是它们可能的排列顺序的一个实例。

表 5.3 学生综合调查的内容

大学图书中心的使用情况、购买习惯和满意度
学术指导系统的使用情况和对这一系统的一些看法
人口学方面的一些问题
酗酒和吸毒:实际情况、对校园有关政策的一些看法、对教育和治疗服务问题的一些思考
校园内种族关系问题

问卷长度与被调查人的负担

为了便于操作管理,电话调查的时间一般在 10～20 分钟,尽管有些电话调查的时间只有 5 分钟,而有些调查时间可能更长一些。个别访谈的时间通常为 30～60 分钟,尽管有些访谈的时间可能长达两个或两个多小时。邮寄式调查的问卷,长度以 4～8 页为宜。当然,无独有偶,有的邮寄式问卷可能稍短,而有的则可能很长。不同的调查方法的问卷长度为什么各不相同呢?一份问卷究竟应该多长?与调查设计的其他方面的决策一样,调查时间的长短不仅需要综合考虑各方面的因素,且常常没有什么明确的法则可资利用。与之有关的因素主要包括经费和其他可资利用的资源、需要收集的数据的数量和类型等。但我们期待的被调查人的兴趣或主动性和调查的实施模式等因素也很重要。

不论可资利用的资源的多寡或我们认为的我们需要的信息量的多少,我们都需要被调查人乐于花费一定的时间完成问卷的填写。我们暂且不谈调查费的支付,而只是考虑怎么才能使被调查人主动地配合我们的研究这一问题。如果研究比较令人感兴趣,那么人们就可能更乐意接受调查。如果它看起来像一个重要的社会问题,那么人们可能会格外的主动。如果在调查之后,根据调查的结果会采取某些行动,那么人们配合调查的主动性可能会因此而大增(Dillman,2000;Groves,1989)。

对于邮寄式调查或其他形式的自填式调查(如电子邮件或互联网调查),问卷的外观形式非常重要,我们必须认真地进行设计。问卷长度的外观形式同样重要,它对被调查人的影响不亚于问卷的实际页数和问题数。迪尔曼(Dillman, 2000)对自填式问卷外观形式的编排问题做了最为完整的介绍。在这里,我们只就其中几个最为重要的方面做一些介绍,对调查—回答模型(survey-response model)做一些考察。

在被调查人决定是否参与调查的时候,他们的心中对调查只有一般的印

象。这时,对被调查人是否参与调查起主导作用的因素是他对负担轻重的感觉。被调查人可能会很快地看一下问卷的版面,包括文字间隔、字型大小和问题的长度等。一个采用小号字型或很长的问题、版面排得很紧的工具往往会使被调查人望而却步。

被调查人可能很快就会将一份长而复杂的答题指南搁置一边。我们希望被调查人在提供我们所需要的信息时,花费的时间和精力尽可能地少一些。含混不清的答题指南也可能会使那些参与我们调查的被调查人的错误回答和题项的无回答现象有所增加。

我们建议(根据迪尔曼的著作)在设计自填式问卷的外观时,大家应该考虑以下几点:

- 将工具的长度限制在6~8页。
- 预先给出每一类答案的编码,即给每一类答案分配一个特定的数码,以便于被调查人打圈。
- 给类别留出足够的空间,以免在给一个答案打圈时触及相邻的类别,类别要垂直排列,而不要水平散布在页面。
- 答题指南要简短,最多不要超过两句。例如"请在您选择的答案前的数码上画一个圈,除非您的答案不在列出的答案中。"
- 问题、答案类别和过渡段落或段落的标题使用的字型要有所不同。
- 只要可能,尽量采用箭头来表明跳答的方位。参见附录A和B。

在问卷设计中应该注意避免的问题

在自填式问卷中普遍存在的两种不甚恰当的倾向是我们应当特别注意的。首先,在一般情况下,应该尽量减少开放式问题的使用。尽管这种形式的问题容易填写,且出现在被调查人将要填写的问卷中显得比较自然,但是它们存在着许多问题。除了前面已经讨论过的那些问题之外,它还会造成其他一些问题,如字迹难以辨认、含混不清的缩写和模棱两可的(调查员无法理解的)答案。这些问题合在一起,可能会给编码和分析带来严重的问题。

其次,不合时宜的、难以理解的词语、行话、俚语经常出现在自填式问卷中,这也许是因为自填式工具(尤其是邮寄式调查)的试调查很少直接在被调查人身上进行,因此无法对它们可能出现的问题进行讨论。此外,诸如这样的调查,一般都用于特定的而非一般的人口总体。我们建议通过以下问题对自填式问卷的用语做一个逼真的检查:

- 词汇是否是日常用语,譬如是否出现在报纸的通栏中?
- 比较简单的言词是否表达了同样的含义?

- 如果该词汇旨在用作某一专业团体的专业用语,那么它是否已为该团体的某些成员所认可(通常情况下,尤其是在物理学之外,即使对技术词汇的理解和使用,也并非完全一致)?

最后,我们选用的试调查方法应该是那些使我们可以从被调查人那里得到一定反馈的方法。得到一份完整的甚至是文理通达的问卷并不能保证这一工具能如我们希望的那样为被调查人所理解和接受。被调查人在回答那些他们并不完全理解的问题时,常常需要揣摩或试答。在有关大学生吸毒政策的问题中,可能大多数被调查的学生可以理解外来语替代(in lieu of)一词,但是有相当数量的学生可能不能理解这一词的确切含义,从而造成答非所问。如果不采用可以收集被调查人的反馈信息的试调查法,这样的问题就不可能被我们发现。

互联网调查也会从这种来自被调查人的反馈中受益。此外,在观察少数被调查人回答互联网调查的问题时,它也很有用处。互联网调查与网上的漫游问题有关。而这种漫游是通过工具、位置信息或完成调查所需要的说明和其他并不适用于传统的纸质自填式工具的有关“可用手段”进行的。这些手段我们将在专门讨论试调查问题的第 6 章做更为详细的讨论。

注　释

1 在电话调查结束的时候,当我们问被调查人他们之所以参加调查的主要原因时,有相当一部分人答道,他们是在不经意间开始接受调查的,实际上并没有认真考虑过是否应该参加调查。

2 对那些需要询问一些比较敏感的问题研究人员来讲,有若干因素是他们必须加以考虑的。有关这方面最新的、简明而全面的讨论,可参见塔伦乔和史密斯在 1996 年出版的著作(Tourangeau and Smith,1996)。

第6章 问卷设计:测试问题

Questionnaire Design: Testing the Questions

试调查是一组测试程序,它是在态度、行为和研究者感兴趣的特征的量度有效和可信的前提下,确定问卷是否能以研究者设想的方式运作。试调查在对问卷进行测试时,测试的对象应该选自将要在主研究中进行调查的某些被调查人。虽然有些试调查方法并不使用被调查人,但是这些方法仍须与基于被调查人的程序配合使用。

尽管人人都认为试调查很重要,但是详细说明它为什么重要仍然是必要的。因为能真正搞清楚为什么我们必须这样做,对完善我们的研究设计是很有用处的。通过讨论不仅会帮助我们更好地理解调查问卷,而且也会帮助我们在众多可资利用的试调查方法中做出明智的选择。

在设计问卷时,我们必须做出许多决定和假设,其中有些我们比较有把握,而有些则不太有把握。构成问卷草稿的核心是我们对有关被调查人了解的事情是什么样的、什么样的词汇能为他们所理解、什么样的信息是他们能够和愿意提供的,以及什么样的答题方式是他们可以操作的等一系列问题所做出的判断。在我们将一组备择选择摆在被调查人面前的时候,实际上我们就开始预想我们得到的数据将会是什么样的了。我们的成就中,有相当一部分源自我们对人们将对我们的提问做何反应的感觉、对自己提出的问题的判断力的自信,以及我们对被调查人生活世界的了解。

在我们的问题的第一稿中,有些假定对于许多被调查人来讲可能是错误的。例如,大多数人是否知道自己每星期开车的总里程数?一般被调查人是否了解出口这一词的含义?大多数被调查人是否能记住他们在过去6个月中一共去了多少次饭馆?上述所有这些问题都有可能成为我们一个调查中使用的题项。不言而喻,被调查人在知识水平、记忆力好坏和参与试答问题的积极性(即使试答并非十分困难的问题)等方面都不尽相同。

在起草问卷第一稿时,我们已经假定绝大多数被调查人将能够和愿意做调查要求他们做的事情——而我们的结果之所以真实可信也正是以这一假定为依据。在起草问卷的时候,要做到不把自己当作被调查人,对于研究者尤其

是研究工作方面的新手来讲,做到这一点是很困难的,更何况这样的情况常常是在不知不觉的情况下发生的。在误把自己当作被调查人时,研究者以为他们自己能够做到的事被调查人也可以做到;他们理解的事情被调查人也理解,甚至认为被调查人也会像他们一样,乐于并认真回答自己认为重要或感兴趣的问题。做出这样的假定是再自然不过的事,它就像人们并未意识到自己做了这样的假设一样自然。只要我们能够确定诸如这样的假定存在,并能对之进行批判性的审视,我们的问卷就有可能有所改进。

在第一次试调查之后,我们就有了一些可以评定其中某些但一般并非全部假定的数据。许多似乎设计得很好的问卷在预测中,一经真正的被调查人试答,便显得漏洞百出。

被调查人对问题的理解和回答问题的能力的重要性

我们在试调查中希望评定的一个问题是被调查人对我们提出的问题提供好的答案的能力。所谓"好的"答案的含义就是有效和可信。效度(validity)首先需要问题能测量我们感兴趣的维度或构念(construct),第二需要被调查人能按我们的本意解读问题;信度(reliability)主要是指在重复试验的答案中存在的变异。例如,假如其他研究者将我们的问卷用于同一总体的同样的调查(在我们调查进行的同时),采用的程序和方法也和我们的调查完全相同,那么那些研究者应当得到(在抽样误差的范围内)与我们相同的结果。

用于确定被调查人对调查的问题的理解程度的试调查方法有若干种。尽管曾经有那么几个研究对这些方法做过比较(Bischoping, 1989; Presser & Blair, 1994; Cannell, Fowler, & Marquis, 1968; Campanelli, Rothgeb, Esposito, & Polivka, 1991; Willis, Trunzo, & Strussman, 1992),但我们仍然无法确定,在某一特定的场合究竟哪一种试调查方法最好。正因为如此,为了最好地配置我们的资源,我们必须自己做判断,而那些有关处理这一问题的带有普遍性的指南,只可用作参考。

用于确定被调查人的理解程度的方法因调查实施的方式而异。我们首先来讨论那些用于调查员填写的数据收集方法(如电话调查或面对面访谈)的试调查法,然后再讨论那些用于被调查人自行填写的工具(如邮寄式调查或互联网调查)的试调查法。

电话调查的试调查需要的主要资源(但不是唯一资源)是人工(主要是访谈的钟点)和可以利用的时间。另外一些资源,如材料、电话费和分析数据所需的时间当然也是必须的,但总费用主要由数据收集所需的人工费构成。试调查不是一个可有可无的预算项目,因此我们在制订调查方案时必须认真考

虑。而没有足够的时间和预算进行周密的试调查，势必导致严重的误差。

常规性试调查和调查员情况报告

在进行常规性试调查时，我们需要调查的数量不多，通常在 20 ~ 40 人的被调查人。样本应该大到足以包括来自目标总体各个子群体的各色人等，这样我们才能对问题和答案类别进行合理的测试。试调查必须完全按照为主研究设计的方法和程序进行。在这样的试调查完成之后，我们应该安排参与该次调查的调查员进行一次任务完成情况的报告会。这样的做法是十分普遍的，一般讲，在大多数人心目中，试调查就是这些。实际上这只是若干种可资利用方法中的一种（Cannell，Oksenberg，Fowler，Kalton，& Bischoping，1989）。

贝斯肖邦（Bischoping，1989）指出，研究者有关试调查的理想的调查员组成的观点不尽相同。有些研究者愿意使用最有经验的调查员（Fowler，1984；Converse & Presser，1986），另一些人则认为试调查的调查员应该由经验程度不等的、各种各样的调查员组成（Demaio，1983），其理由是缺乏经验的调查员有可能发现那些有经验的调查员几乎可以下意识处理的问题中的问题。我们则比较倾向于前者。在可控实验中，贝斯肖邦（Bischoping ，1989）发现，在他们提及一些或某些类型的问题中，调查员经验水平不同调查结果也显著不同。

在试调查开始之前，做好两件事情是很有用的：第一，我们应当完全像计划主研究那样，对试调查的调查员进行培训。特别要注意的是，我们应当一题一题地（Q-by-Qs）给他们提供填答说明。如果我们希望调查员回答被调查人提出的有关某一词汇的含义的问题，如“医务人员”或“刑事犯罪司法系统”，就需要在试调查中对他们进行这方面的培训。这些一题一题的说明将会对被调查人的回答产生影响。当然，在小规模的试调查中，可能无法遇到某些在主调查进行时可能会遇到的在回答问题用词方面的麻烦。例如，只有一小部分被调查人无法确定“医务人员”一词的确切含义（在调查的场合中），那么一个小规模的试调查也许就无法发现这一问题（这正是多次试调查法之所以有用的原因之一）。

第二，调查员以前没有做过试调查，那么事先花费一点时间来讨论试调查的目的意义和我们需要调查员注意什么样的情况则不失为一个好主意。如果调查员了解我们的兴趣在于被调查人是否理解某一词汇的含义，或被调查人是否能记住答案的类别，那么便会有助于他们对被调查人的有关行为予以特别的关注，并进而将有用的信息记在情况报告书中。教会调查员如何在每一个试调查之后在问卷上做一些记录，是使情况报告书的内容更丰富和更精确的一种简单方法。

但是我们必须提醒大家:我们可能会在不经意间在调查员心中“植入”了某些问题,而这些问题会被调查员认为很重要,从而在情况报告书中反复出现。假如我们对调查员说:“我认为有些人在考虑如何回答‘您将有多大可能在(自己的)住宅被抢劫或被谋杀’这一问题时,的确会有一些问题”,那么调查员就有可能过分报告被调查人在回答这一问题的难度。甚至有可能因为先入为主地认为问题很困难,有些调查员可能不给被调查人足够的时间来考虑自己的答案,或更乐于被调查人选择回答“不知道”。这种做法会影响某些被调查人的答案,进而影响我们的数据。正因为如此,我们建议在填写说明中,有关调查员的注意事项那一部分的内容应该更为普遍和全面一些。

调查员情况报告书的内容和结构一般如下:

- 试调查情况概述,内容包括任何有关导致被调查者拒绝参与调查或拒绝回答某些调查主题的严重问题。
- 一题一题确定每一题项存在的问题,必须询问所有的调查员。首先问调查员发现的有问题的题项,然后再问被调查人在回答每一有问题的题项时出现的所有问题。
- 一题一题地提出修改意见,这时调查员可就已经确定的问题提出修改意见。
- 总结评议,内容包括对已经完成的试调查的优缺点进行总结和在实际进行数据收集或再次进行试调查前应注意解决的问题。

注意,尽管我们感兴趣的是被调查人的理解能力(以及其他与被调查人有关的问题),但是情况报告还是应当包括其他一系列问题的报告和调查员自己的观点。这一点是很重要的,我们必须牢记在心。实际上,调查员的作用好比试调查中被调查人的代理人。在做情况报告时,区分调查员考虑到的被调查人可能会发生的问题和被调查人实际发生的问题这二者是很有用处的。

此外,在一个情况报告中,得到所有调查员对被调查人的经验和反应的评论也是很重要的。情况报告常常可能会为一两个善于言词的调查员所主导,而其他调查员则没有机会报告他们掌握的有用信息,发表他们对试调查的评论。为了解决这一问题,我们可以先询问每一个到会的调查员,然后再对问卷的每一题项进行评论。

我们请大家注意的另一个问题也许可以叫做奇闻轶事。我们在调查中遇到的极端情况往往令人难以忘怀,因而几乎总是会在情况报告中报告。这方面两个比较普遍的例子是知识异常渊博的被调查人和置身事外的被调查人。第一种情况,譬如在我们的犯罪调查中,一个调查员可能会调查到一位律师,他在对许多调查问题提出的技巧颇为赞赏的同时,不仅给我们提出了一些备择的答案类别,而且还提出了许多问卷中未曾包括的问题。至于第二种情况,我们还是用同一个研究例子来说明。调查员有可能遇到一位十分健谈的被调

查人,他历数了他的家庭成员经受过的各种犯罪行为,这不但使调查时间过长,也许还会令人心生疑窦——这位被调查人可能不愿谈及自己家庭成员的犯罪行为,对有关这方面的问题高度警惕。这样的情况可能会增加情况报告会的时间,如果这样的情况发生在一个"多嘴"的调查员身上,情况就尤其如此。在我们从两个被调查人身上了解到一些情况的同时,我们必须时刻牢记这个被调查人是一个例外,我们在对问卷进行修改的时候,不可被他们所左右。当然,大多数被调查人既不是法律专家,也不是经常受到犯罪行为侵害的人。

另外一方面,试调查还告诉我们,被调查人常常会拒绝回答那些他们认为是不真实的问题。我们来看一下有关判决的问题:

您认为犯有抢劫罪的人应该入狱,还是应该要求他们向假释官每月报告一次、每天报告一次,或受到电子监控,以便我们随时都可以知道他们身在何处?

虽然许多被调查人,甚至是大多数被调查人也许乐于回答这种形式的问题,但是另一些被调查人则可能不愿意,他们的答案可能取决于犯罪的具体情况。他们希望知道抢劫是否使用了手枪、是否有人受伤、抢劫犯是否有前科等情况。如果这一问题对我们的研究至关重要,而我们可能是从其他研究或探索性研究中知道影响被调查人的选择的关键因素,我们可以选择采用一种更为复杂的问题形式。在这一例子中,我们可以在要求被调查人在备择的观点中做出选择之前,提供一幅小插图[1]。例如:

一个[20]岁抢劫犯在街上抢了100美金。他用[枪]威胁受害者。他有[4]个前科。

然后我们再问被调查人,他认为应该怎么处理这个罪犯。方括弧中的内容随调查(即访谈)对象变化而变化。但变化遵循的模式是使最后生成的数据集能使我们得以分析被调查人对不同犯罪情况所持有的不同观点。不言而喻,这种形式的问题无论是在调查的时候,还是在分析的时候,都比前一种形式要复杂得多。但是它却更加真实,信息量也更大。这样的问题可以使我们了解更多的情况,但我们也必须为此付出一定的代价。

常规性试调查是一种强有力的工具,但要把它用好却不是一件容易的事。我们已经注意到了在情况报告中某些可能发生的问题。情况报告会的主持人必须做好充分的准备,并十分有技巧地掌握好会议。不仅如此,有时情况报告会上的某些观点可能彼此冲突,例如有些调查员可能认为某一个问题存在一定的问题,需要做进一步修改,但另一些调查员却认为它很好。有时,有些较有经验或头脑清楚的调查员可能非常自信,即使实际试调查的结果并不支持他的观点。常规试调查长于发现可能存在的问题,但却无法为我们提供解决

问题的线索。

调查后的调查

另一种得到有关被调查人理解信息所采用的比较普遍的方法是和被调查人讨论刚刚做完的调查。这种方法以贝尔森(Belson,1981)的著作为根据。在使用这种方法时,不同的调查员将再一次与某一调查的调查对象进行联系,并向他们询问第一次调查时题项的含义。贝尔森发现许多被调查人并不了解原来调查问题的本意。目前采用的实际做法是在调查按规定的要求完成之后,接着立即询问被调查人后调查问题。这时我们必须使后调查问题尽可能简短,因为被调查人刚刚花了不少时间完成我们的主调查。

调查员可以问类似这样的问题:"现在我想了解一些有关我们刚刚做完的调查的情况。"一般讲,我们希望能用事先准备好的后调查问题,但是不要仅限于这样的问题。如果我们从调查中得知被调查人对某些题项的理解存在问题或混乱,这时是否要将这些问题包括在后调查中可由调查员自行决定。

下面是一个后调查问题的实例:

这些就是我想正式问您的有关犯罪的所有问题。现在我想问您几个有关调查的问题。

1. 第一个问题是,有没有什么您不知道怎么回答的问题?

如果有,是哪几个?

为什么您不知道怎样回答[问题的题目]这一问题?

2. 我在使用刑事犯罪司法系统一词时,您知道它的含义是什么吗?

3. 在我问您有关您的邻里地区的性质的问题时,您考虑的是什么样的事情?

4. 您是否认为有什么问题大多数人都难以回答?

如果有,这些问题是哪几个?

为什么您会认为人们在回答有关[问题的题目]时会有困难?

5. 您认为还有什么重要的事情没有包含在我们的问卷中?

注意,与问卷中的问题不同,后调查题项大部分都是开放式的。因为我们无法确定被调查人可能指出的问题是什么类型的,所以我们不可能过分限制他们给出的答案。如果这是系列试调查的一部分,那么在第一次试调查中确定的那些问题,可以作为随后试调查中加以明确的问题。

例如,学生综合调查中前面的后调查可能会揭示诸如这样的事情,即学习指导的易得性对学生的重要性。不仅如此,同辈的指导老师,在班内的排名至少不低于他们所指导的学生。诸如这样的事情,研究者因为并非调查总体的

一员，往往不甚了解，但是诸如这样的认识，只要研究者允许参与试调查的被调查人对调查的问题进行评论就能很快得到（这一点也正是焦点组访谈所强调的）。

有两种途径可令被调查人参与诸如这样的讨论。第一种途径是在调查开始的时候，我们就告诉他们，在调查结束之后我们将要和他们讨论有关调查的问题。坎福斯和普雷舍（Converse & Presser，1986）把它称为参与式试调查（participating pretest）。还有一种途径是先按计划进行调查，然后再问有关后调查题项。这种方法叫做未宣式试调查（undeclared pretest）。坎福斯和普雷舍建议我们在多次试调查中，开始先采用参与式试调查，然后再使用未宣式试调查。但是论述将两种方法结合起来使用的效果的文献可谓凤毛麟角。

行为编码

行为或互动编码法系卡内尔等人（Cannell et al.，1968）在密西根大学提出的。这种方法旨在从数量上评定面对面或电话调查的质量。在这种方法中，我们假定在理想的调查员和被调查对象的互动中，调查员逐字宣读问题，而被调查人则以可以接受的答案回答问题。在处理完一个问题之后，再进行下一个问题。在互动编码时，第三方使用一组简单的代码指出对于这种模型的背离：没有逐字逐句宣读问题、被调查人要求澄清问题、要求再次宣读问题，或答案不够精确。每次每个问题都要进行编码，每种类型问题的编码，如果没有发生为0，否则就为1。然后分别计算每个问题的每种类型问题发生次数的百分比。

表6.1 标准行为编码

调查员
E（准确）
S（轻微背离）
V（确认）
WV（错误确认）
M（严重背离）
被调查人
AA（回答准确）
IN（回答中断）
QA（经确认的答案）
CL（要求解释）
IA（答案不准确）
DK（不知道）
RE（拒绝回答）

尽管这种方法的本来目的是调查调查员的调查质量，但它也可用来确定有问题的问题。福勒（Fowler，1989）建议把那些某一类背离编码在15%以上的问题作为有问题的问题处理。虽然以15%作为标准多少有些武断，但是在实际的调查设计中，把它作为一种参考，则不无用处。这种方法的一个优点是在进行电话调查的常规试调查中，我们只要很少一点费用就可以在集中的电话设备中，使用监控设备做这样的编码。而对于个别访谈而言，则一般需要先进行录音，然后再进行编码。表6.1是常用的行为编码。

互动背离告诉我们有关问卷的什么呢？互动问题可以因为调查员的操作不当或任务过于困难、被调查人方面的某些问题或问题表达本身的问题而产生。究竟是什么引起一个特定的问题，这一点并不总是很清楚的。在存在一定的疑问时，我们应该将疑问继续留在问卷中，还是把它从问卷中清除出去？

违背逐字逐句地宣读问题这一要求的原因可能是因为问题措辞不当或问题过长。这一背离可能是因为调查员虽然力图照本宣科地宣读问题，但问题本身存在的种种弊端使他难以做到。然而诸如这样的背离也可能是因为调查员有意改变问题的措辞所致。

在宣读问题时做的轻微变动是指那些完全没有改变问题本意的变动。这样的变动可以包括诸如增加"补白"词或短语这样的修改，例如增加像"……下一个要问的问题是"（接着便准确地逐字宣读题项）的补白。诸如以全程替代缩写这样的变动也属于轻微变动。更改一个词，或删除或增加一些词则被视为严重变动，因为这样的更改改变了问题的句子结构。

在被调查人方面，中断回答可以理解为问题的措辞不太恰当，因误导被调查人而导致问题终止。一个被调查人也可能因为根据他听到的问题的那部分内容知道（或他自以为知道）问题的答案而中断一个问题。这样的中断之所以发生只不过是因为被调查人回答得太快。但是，中断的发生也可能因为被调查人对问题感到迷惑不解而引起。中断可能因为问题存在相当小的结构性问题而引起，但它也可能意味着存在某些更为严重的问题。

当被调查人对自己答案加以确认时，这就意味着他们不清楚问题的含义，因而不知道自己的回答是否正确。但是有时，被调查人之所以对自己的答案加以确认是因为他们觉得在回答问题时存在一定的困难，因而无法确定自己的回答是否正确。这样的情况大多发生在回答一个需要进行艰苦回忆的问题时。一个不够精确的答案也可能因答题的任务过于困难，使被调查人知难而退所致。

要求做进一步的解释通常是问题存在某些问题的可靠信号。一个好问题是无需任何解释或任何额外的信息便可理解。从这几个为数不多的例子中，我们不难看出，从行为表现的蛛丝马迹来判断行为的原因不是一件容易的事。正因为如此，尽管在一个互动背离经常出现时，我们可能因此而觉察到有什么出现了差错，但我们却不能就此而确定引起背离的确切原因。

在表 6.1 列出的基本编码之外，有时我们还会使用一些其他的编码。例如，我们也许想了解，为了得到精确的答案，调查员会经常做一些什么样的探讨、他们在向被调查人提问的时候是否按照培训时提出的要求做或被调查人是否经常更改自己的答案等问题。虽然我们可能有许多条充分的理由来使用其他额外的编码，但是我们还是要奉劝初涉调查研究的研究者，要学会抵制增加使用其他编码的诱惑。

但是,如果存在不可辩驳的使用额外的编码的理由,那么它们应该满足一些什么样的标准呢?首先,它们应当能捕捉到重要的信息。研究者必须清楚额外的数据应该如何使用。如果它们只是一些有兴趣的问题,那么就像我们在问卷中加入额外的问题所应遵循的原则那样,我们应该避免加入这样的额外编码。第二,它们必须是定义明确的。这就是说,它们不能和其他编码有任何重叠。第三,必须对辅助性编码予以清楚的说明,以便编码员能够可靠地使用它们进行编码。最后一点,尽管我们也可以把这一点放在前面,但是我们觉得还是放在最后更好,因为对这一点的看法似乎仍有争议。但对于辅助性行为的发生应该予以必要的关注这一点似乎不存在什么异议。

在加入过多的额外编码时,15%的限制这一标准通过的可能性将会随之加大。正因为如此,这就有可能在某种程度上使被标出有问题的问题数成为使用的编码数的函数。

行为编码的分析经常被限于对于每一问题背离的总数进行简单的总体性描述。在有些场合,我们的分析也许可以更加深入一些。例如,我们可以用它们来做不同调查员的背离码频数分析。由大量背离引起的异常值可能会对结果有所影响。而这一异常值是很容易发现的。如果试调查的规模比较大,那么做一些其他类型的分析也是可能的。背离是否可能由调查员的操作而非工具本身引起,且随时间推移而有所或没有下降?如果在做了许多调查之后,误读有所下降,那么调查员的培训和实际操作,比问题的措辞对降低背离更重要。

行为编码的主要优点在于费用低廉、概念简单和数量化。编码员可以按照计划进行相当快速的培训。行为编码很容易和其他试调查法结合起来使用。监控常规试调查的督导可以对这些调查的行为码进行检查(此处原文似乎在 can 后面漏掉一个动词。——译者注)。诚如福勒(Fowler,1995)所言,在将行为编码用于调查员对被调查人的问题做主观判断时,我们可能还会有额外收获。如果一个常规试调查的情况报告中说,调查员发现许多被调查人在回答某一特定问题时都存在一定困难,那么我们也许可以有望在行为编码的数据中发现某些证据。

行为编码法的不足在于,他不能提供问题之所以有问题的信息。作为一个能为我们找到存在潜在问题的、行之有效的方法,它为我们提供的可用于改进我们的问题的信息却太少。正因为如此,在试调查中只使用行为编码一种方法是远远不够的。

跨编码员信度问题

不同的编码员在使用行为编码时,其一致性程度如何?理论上讲,如果两

个编码员听取同一调查员——被调查人的互动,那么他们给出的编码应该是相同的。有关这一信度的一个不言而喻的量度是:两个编码员给出的相同编码次数的百分比。一致性次数百分比的不足之处在于,存在编码员的一致是偶然所致的可能性。一种更好的编码员一致性量度是科恩的卡帕系数(Cohen's kappa),这一统计量对诸如这样的偶然一致做了修正。有关科恩的卡帕系数的详细讨论超出了本书讨论的范围。希望对这一问题有更进一步理解的读者可参见科恩(Cohen , 1960)和埃福瑞特和海(Everitt & Hay, 1992)的有关著作。在条件允许的时候,在编码员培训完成之后,让所有的编码员对同一个调查的录音带进行编码,以在真正的编码开始前评定跨编码员信度。

认知访谈

用以确定被调查人理解程度的另一种试调查法是做一对一(one-on-one)或认知访谈(cognitive interview)。在这种参与式安排中,被调查人是征募来的,并要求他们处于访谈的关键位置。问卷是被调查人单独进行填写的。我们从认知心理学那里借来一种叫做边想边说(think aloud)的方法,被调查人在回答每一个问题时,根据给出的指示做出声的思维。其目的在于了解被调查人在回答问题时心中在想什么。这种方法使我们能够像其他试调查法那样深入洞察理解问题,但采用的方式更为直接,因为被调查人在回答一个问题时,清楚地报告他们在想些什么。除了帮助理解之外,这种方法也可以帮助我们确定回答过程中的其他阶段的问题,如帮助我们确定执行诸如回忆问题、使用答案的选项等任务中发生的问题。

我们给被调查人提供的指示形式如下[基本上以艾利克松和西蒙(Ericsson & Simon, 1993)的文本为准]:

> 我想问您一些问题,我们提问题的目的在于了解您在回答问题的时候心中在想什么。为了能达到这一目的,我们希望您能边想边说。所谓边想边说就是我们希望您能告诉我们,在从开始听到我们提出的问题起,到给出答案为止的整个过程中,您想到的所有一切事情。我们不希望您周密地考虑您所说的,或对我们解释您给出的答案。您的行动就像您独自在一个房间里自言自语。最重要的是您要不断地说。如果您长时间沉默不语,我将要求您开口说话。

调查员将必须偶尔提醒被调查人保持谈论,这时他可能需要使用诸如“您在想什么?”或“您会给出什么样的答案?”这样的探询语句。一般讲,在访谈过程中,在被调查人给出问题答案之后,调查员也会立即提出一组事先准备好的,即照本宣科地提出一组探询性问题。虽然后一种做法与认知心理学家

使用的旨在引出口头报告方法有所不同，但是它却为调查研究者所广泛使用。由此而产生的方法是边想边说和贝尔森式的“深度访谈（intensive interview）（Belson，1981）”的混合物。

边想边说经常作为问题指向式探询（question-specific probes）的补充。实际上，实际调查研究者对于诸如这样的探询法的依赖更甚于边想边说（随后还会依赖得更多）。探询是一些用来了解被调查人对问题的某些方面的看法的问题。例如，一个探询问题可能问道“您怎么理解问题中‘亚人种’（racial profiling）一词的含义”；或者更为宽泛地问道“能否以您自己的语言表达这一问题的含义？”探询性问题既可以是事先为访谈准备好的书面形式的，也可以是调查员在认知调查过程根据实际情形编造的。

康拉德和布莱尔（Conrad & Blair, 2004）发现，我们固然可以卓有成效地教导缺乏经验的调查员学会使用边想边说法，然而那些更多地依赖于非书面的即兴探询技巧需更有经验的调查员。除非我们以前有过做认知访谈员的实际经验，否则最好还是使用书面的探询性问题。如果我们事先未能确定探询性问题，那么一个缺乏经验的调查员很容易使被调查人“确定”一些实际上并不存在的问题，或只是进行一些毫无效果的探询。在为数不多的用于认知访谈培训的资料中，有一本是威利斯著的《认知访谈和问卷设计：调查员培训手册》（*Cognitive Interviewing and Questionnaire Design: A Training Manual*）（Willis，1994）。

在了解被调查人是怎样形成自己对一个问题的答案的过程中，我们得以深入了解他们是如何理解问题的。例如，在回答“去年您是否经常喝酒？您觉得您每星期喝一次以上、每星期喝一次、每个月喝一或二次，还是一个月喝不到一次”这一问题时，一个被调查人边想边说，就可能提到诸如这样的事：“去年我喝得不太多。除夕夜有一个聚会，当时可能是在感恩节前的最后一次，那时一伙人聚在一起观看巴黎圣母院游戏（the Notre Dame game）……”在我们检查边想边说的抄本［称作记录（protocol）］时，研究者认识到被调查人显然是在考虑去年，而非调查前的12个月。如果后者正是我们所要求的，那么这一问题也许应该改为“自从您上次（给出一年中早于调查进行的那个日子的一个具体日子）喝酒以来的12个月内，您是否喝过任何一种酒？”另一种可供我们选择的方法是用更清楚的探询问题，如“您在回答这一问题时，您想到的时间是过去的12个月，还是去年？”直接询问被调查人。诸如这样的探询可能更有效。如果边想边说没有发现发生在调查涉及的时间内的问题，那么诸如这样的探询可能浪费我们的时间。更为重要的是，在编造探询问题时，我们必须倍加小心，以免它们指出的问题对于被调查人来说并不存在。恰如常规访谈的访谈员探询一样，试调查的访谈员探询也可能发生非我们所愿的效应，这样的情况一旦发生，就有可能使认知访谈的效应发生偏倚。正因为如此，我

们奉劝那些调查研究方面的新手,最好不要在访谈的同时编造探询性问题,而要使用书面的探询性问题,这样便可以事先仔细地检查是否可能存在某种偏倚。

在被调查人面对的任务或问题的概念有一定难度的时候,认知访谈尤其有用。有关这些方法完整的讨论和它们的理论依据可参见爱立信和西蒙(Ericsson & Simon, 1993)的著作。有关它们在调查中应用的实例则可参见路易斯通(Royston , 1989)的著作。有关认知方法在测试调查问题和研究应答过程中使用的文献正在迅速地涌现(参见 Schwarz & Sudman,1996;Schwarz & Wellens,1997;Sudman,Bradburn & Schwarz,1996;Tourangeau,Rips, & Rasinski,2000;Conrad & Blair,2004)。

认知访谈的另一个优点是可以快捷地发现许多问题。然而因为认知访谈使用的样本一般都很小,所以我们难以由此了解被调查人可能遇到某一特定困难的频度。当然,如果时间和条件允许,我们可进行样本较大的试调查,但是有关最佳样本量的研究并不多见。更为重要的是有关对大量认知访谈的结果进行定量分析的方法的文献更是凤毛麟角。

与后访谈法(postinterview method)相比。在认知访谈中,被调查人在回答问题时的思路显得比较清晰。它的不足之处在于这种类型的访谈需要动用的资源比较多。我们需要做更多的联系工作,以找到愿意来接受我们调查的被调查人。一般讲,一对一的调查与一般的电话调查不同,我们必须付给被调查人一定的报酬。报酬一般在 10 ~ 50 美元不等,具体多少取决于调查需要的时间的长短。我们提醒大家注意的最后一点是,认知访谈一般都由年资比较高的调查员实施。这样的调查员不仅人数有限,且酬劳也比较高。正因为如此,在预算固定的前提下,我们最终都选择做较少的这样的调查,而更多的做突击式的电话调查。

对认知访谈得到的结果进行的分析经常都是很不正式的(Tourangeau et al. , 2000)。在很多情况下,调查员在仔细检查自己在调查期间做的笔记时,好比一个也在做笔记的观察员。根据这些观察,他们确定哪些问题是有问题的。通常他们考虑的是问题的性质,而有时他们也会考虑怎样改进的问题。在诸如这样的过程中,只要稍不留意,他们的考虑就会超出调查资料的范围,从而使这样的考虑以他们自己对问题的判断而非被调查人的任何意见和实际行为为依据。

康拉德和布莱尔(Conrad & Blair, 1996, 2004)发现,在独立的分析员试图根据认知访谈的抄本确定某些特定的问题是否会给被调查人带来问题时,其得出的结果并不十分可靠。这就是说,如果两个(或更多个)分析员对同一抄本进行分析,他们对问卷中的问题是否存在问题得出的结论不尽相同。尽管存在这样或那样的不足,但是近 20 年来,各种各样的认知访谈法还是一直

在被广泛地使用着。

我们一直把讨论的重点放在认知访谈怎样揭示理解中存在的问题上。然而,仍然存在另一种可能,即一个完全能理解我们的问题的被调查人,在执行回答问题所要求的其他任务时却有很大的困难。例如,被调查人可能不难理解"去年您看了多少电影?"这样的问题。但是,如果一个被调查人去年看了很多电影,那么对他来讲回忆和计数看过的电影就相当困难,当然也因此而很容易出差错。虽然常规性试调查有可能比较好地揭示这种困难,但是采用边想边说的认知访谈则能更好地揭示诸如这样的困难。不仅如此,在我们聆听被调查人的边想边说时,某些任务对于某些人来讲有一定困难的简单事实的详细信息可能会出现。

在一个由西摩·苏德曼主持的研究项目中,主要采用边想边说的认知访谈显示,这种方法对于了解被调查人是怎样回答有关行为频数的问题是十分有效的,且对自我报告和代理报告(self-reports and proxy reports)二者都同样有效(参见 Blair, Menon, & Bickart, 1991)。

虽然认知访谈在大多数情况下都被用于由调查员填写的调查,但是只要稍加变通,认知访谈技术就可以用于自填式工具的测试,无论它是常规性邮寄式工具还是互联网调查的问卷(Schechter & Blair, 2001)。认知方法在自填式工具方面的使用主要是组织的调查(Willimack, Lyberg, Martin, Japec, & Whitridge, 2004)。此外,在一个为美国人口普查局做的、调查种族和少数民族问题中出现的新问题的项目的一系列研究中,进行了很多的认知访谈(参见 Davis, Blair, Gourdreau, Boone, Johnson, & Robles, 1998)。

认知访谈对回答问题的影响可以发生在任何一种填写模式的调查中,但是它对自填式工具的影响可能更大。在回答一个书面的问题时,被调查人必须在阅读工具的同时(一般自行阅读)进行交谈。绝大多数自填式工具都含有供被调查人使用的填写说明。其中也必须包括定义和逻辑跳跃的说明。而这些加于被调查人多重的要求势必降低他们执行调查任务的质量,从而导致错误的发生和任务难度的增加。如果没有这样的多重任务,诸如这样的问题就可能不会发生。

认知访谈法的一个重要作用是最大程度地减少"中断"影响。对于自填式问卷而言,做到这一点可以较少地依赖实时的边想边说,而更多地依靠回顾性的边想边说。在做回顾性的边想边说时,被调查人是在回答之后,而非在回答问题的同时报告自己的想法。在工具自然中断的地方——页脚、段落间都应当用探询性问题做补救。

总之,对于初学者我们建议他们在工具构建的早期使用认知访谈时,最好以边想边说法和以使用书面的探询性问题为主。

被调查人情况报告

被调查人情况报告与后访谈的访谈略有不同。在这种情况下，在完成访谈之后，被调查人被一系列问题带回整个访谈，按每一题项逐项进行探询。这些情况报告通常都是在面对面的情况下完成的，好比将被调查人带到意见中心办公室进行询问那样。最初的访谈一般做了录音（或录像）。录音带可以进行回放，以帮助被调查人回忆他们在实际回答问题的时候心中在想什么。这样的情况报告所需的费用与认知访谈的相仿。区别在于，它可以使用水平较低的调查员，但却需要一小笔其他的费用，购买录音设备和消耗品。

我们应当提醒大家，所有这些特殊的方法都需要在对调查员进行实际的调查之外进行特殊的培训。不仅如此，在联合使用的方法的个数递增的同时，虽然额外的费用也在同步递增，但有关什么样的联合才是最优联合的论著却未见递增，这方面的文献至今仍不多见。对于那些有许多新的或可能的困难题项的研究来讲，我们建议刚开始时使用一种便于被调查人理解和直接清晰地回忆的探询式问题的方法，如后访谈法或认识访谈，然后用行为编码过渡到常规性试调查。对于非常简单的工具，常规性试调查可以单独使用。

专家组

最后，普莱泽和布莱尔（Presser & Blair, 1994）发现，在确定问卷中存在的问题时，使用专家组这种方法比使用其他的方法更有效。专家组是一小群被请到一起对问卷进行评议的专家。专家组的成员主要是调查的专家，也包括个别研究主题方面的专家。专家的人数通常在 3 ~8 位不等，在评议会前的若干天，我们给每一位专家分发一份编写好的工具的草稿。然后再进行小组讨论会，与会的专家对题项逐一进行评议，这一过程与常规的调查员情况报告相似。区别在于专家们的意见来自他们对问卷（或所涉专题）的知识，而非试调查的被调查人的反应。普莱泽和布莱尔（Presser & Blair,1994）发现，这样的专家组不仅在确定其他试调查法发现的许多类型的问题时是很有效的，而且在发现其他方法发现不了的问卷中存在的瑕疵（如隐性的分析困难）也很有效。专家组的能力来自他们掌握的各种各样的专业知识和在会议上彼此的交流。普莱泽和布莱尔的专家组由一个认知心理学家、一个问卷设计专家和一个调查方面的方法学通才组成。当然，我们也应当尝试一下其他的组成方式。重要的问题是我们必须注意到，有时，无论在对潜在的问题，还是在对解决问

题的方法的认识上，专家组未必都能达成共识。其实其他方法也同样如此。我们不能因此而反对使用专家组。而我们应当明确反对的是用专家组替代对被调查人做试调查这样的错误做法。

即使是手头拮据，支付不了与会专家报酬的调查研究方面的新手，仍然有机会采用专家组法。因为专家组会议需要的时间并不多，通常总共只需三四小时，因此我们经常有可能找到一些专家，请他们参加一两次会而不付任何报酬。大学的教员也许可以是我们物色的对象（特别是学校的教学项目），其他大学的同事可能也可以成为我们物色的对象。我们还可以物色的对象可能是附近地区的调查组织内的专业人士。最后，那些曾经在我们研究题目上做过研究的研究者也可能对我们有所帮助。（这些最后提到的、可能成为专家组成员的人，可以在项目实施的文献检索阶段加以确认。）如果差旅费和时间有限，我们可以设法通过电话进行专家组会议。这样一种形式的会议，费用比较便宜，但是我们认为，专家组会议的首选形式还是坐在一起可以当面进行交流的会议。

评定调查员的工作

调查员工作的好坏对调查质量的好坏至关重要。诚如我们所知，调查员可能成为各种调查误差的主要来源。幸运的是，在资源有限的前提下，许多揭示来自被调查人的误差的方法，对确定调查员造成的误差和问题也同样有效。

不言而喻，调查员情况报告是调查员报告自己在调查中遇到的困难，如令人费解的问题、逻辑跳跃模式中的逻辑错误、难以发音的词汇，或朗读和刊印的文字不同等问题的良机。实际上，我们必须加倍小心，在情况报告会中，不要让过于详细的调查员的问题报告掩盖了被调查人的困难报告。

监控电话调查的试调查（和主调查）的调查在大多数调查组织中都是一种标准程序。电话监控设备允许一个督导和调查负责人，在调查员和被调查人毫无知觉的情况下，在调查进行的过程中监听调查。在正式的数据收集过程中，这一措施用来对调查的质量进行控制，了解问题的问答实际过程是否符合规定的程序。同时，这也是一种在试调查中测试调查中存在的问题的好方法。我们力劝研究者对实际的访谈过程进行监控。有时听和读出现的问题大不相同。那些在纸面上看似有一点别扭或复杂的语句，在调查中常常是令人难以忍受的。对于调查研究方面的新手而言，监控也对他们学习如何编写听起来十分自然的问题很有帮助。

修改和再测试：确定什么样的试调查问题需要解决

解决试调查发现的每一个问题的想法既不可能，也不明智。首先，我们也许可以断定一个确定了的"问题"实际上可能并不是一个问题，或是微不足道的，所以不必为此劳神费钱做什么修改和修改所要求的额外的测试。调查人员的烦恼——通常出现在试调查的情况报告中——并没有在问卷的下一稿中得到解决，这样的现象是很普遍的。无独有偶，有时我们可能会发现，有些问题是很少发生的，如果我们把这样的问题也考虑进来，就可能引起其他的问题。例如我们假定试调查的问题是"在喝酒的大多数时候，您平均一次喝多少？"某一个被调查人说道，当她和男朋友一起出去的时候，她总是从男朋友的杯中喝一两口，而不是自己单要一杯。这样，她就很难回答"喝多少"的问题。而为了像这样的被调查人也能够回答这样的问题，我们势必要对原来的问题做很大的修改，而这样的修改会使问题变得使大多数的被调查人感到太复杂。在这样的情况下，也许明智之举是保留原来的问题，不要做任何修改。

除了这些情况之外，我们建议大家在考虑如何对问题进行修改的时候，参考一下表 4.1 列出的问题效用指南，这对我们的修改工作将不无帮助。如果根据这些标准，试调查发现的问题影响了问题使用，那么这一问题或需要进行修改，或干脆删除。有时，我们发现某一问题只是难以发问，其原因则可能是研究者的目标过高过复杂，需要把一个问题分成多个题项。如果问题是我们研究的关键之所在，那么明智的做法是把它拆分成多个题项。如果问题不是很重要，我们就可能会认为那些需要增加的问题是不值得的。在另一些情况下，问题也许不适合用于我们采用的调查方式：问题可能过长或有过多的答案类别，不适合在电话调查中使用，在这样的情况下，如果问题不能得到有效的修改，明智的做法是将它删除。

确定需要做多少次测试

我们已经介绍给大家的各种不同的测试方法，显然可能是相当精细的、费时和费钱的。我们应当预先考虑到试调查需要的工作量，并根据工具的复杂性、新题项的比例，同时适当地考虑到逻辑跳跃模式结构问题做出一个计划。如果有很多部分只是问一部分被调查人，那么我们可能需要做多次试调查，以确保问卷的所有部分都能被测试到。在完成了所有的测试之后，我们确信我们可以用问题效用指南对试调查的结果进行评估，并且的确做出了评估，那么

我们就必须做出困难的选择，或删除无用的题项，或从项目的其他部分挪用一些资源（包括时间和资金），进行进一步的测试。这最后一步只有在两个条件都具备的时候才可以实施：①有问题工具题项是我们研究的问题的关键；②不会对其他部分的设计（如主样本量）产生重大影响，引起新的问题。

试点调查

最后一种用于设计和改进工具的方法是试点调查。在试点调查中，我们采用严格按照将来准备在研究中使用的程序和方法进行相当数量的调查。在进行试点调查之前，我们首先要进行若干试调查。诚如我们所知，在试调查中，我们用少量的个案来探测问卷中的各种不足。我们试图改进工具和调查的方法步骤，以最大可能地争取被调查人的合作，并帮助他们完成我们要求他们完成的任务。但这些方面的问题主要都是数据收集阶段的问题，并非只靠试调查就能发现解决。

例如，试调查的样本一般都过小，无法用于对合作率或合格的被调查人筛选率的估计。此外，某些现场使用的方法步骤可能无法用小样本进行测试。可能存在某些未知特征的样本框，或目标总体对使用的工具的某些方面的应答特征。这样的特征只能通过大规模的现场测试才能检测到。那些对一个小的但是很重要的子群体会有影响的问题，无法用一般试调查使用的小样本探测到。

当主研究的费用相当高，或某些方法步骤是以前未曾采用过或非同一般的时候，那么不做试点调查就进行正式调查可能会有较大的风险。总的讲，试点调查的费用应该是可以承受的，但这只是在进行了充分的试调查之后。在试点调查之后再做重大的修改是非常低效的。

因为大多数小规模的调查都无法进行试点调查，因此最好使用那些已经充分测试的实地调查和抽样方法。为了能达到这一目的，在研究工作开始的时候，花一点时间，像我们前面已经讲过的那样，从调查问题和在我们研究的目标总体上已经做过的调查两个方面进行文献检索是很有用。数据收集方式是否适用、是否能得到被调查人的配合、是否存在数据质量问题和那些曾经使用过的、特定的调查员培训方法（如怎样处理敏感问题）是否适合我们的调查等，这些问题都应当在考虑之列。

尽管有关怎样使这些方法最优地联合起来，以最小的成本，使我们从试调查所发现的问题得到最大程度地解决的理论或实际研究为数不多，但是将多种方法合在一起使用的做法已经为人们所普遍接受。

若干最后的忠告

问卷是我们进行研究所不可或缺的工具。正是通过问卷,我们才得以将被调查人的观点、行为和特征转化为数据。第 4 章和第 6 章的重点是设计改进和测试问卷,并说明工具在调查过程中所处的地位是举足轻重的。不仅如此,我们可以讨论的问题还有许多,因而在那些读者试图构建问卷的第一稿的试调查中,我们可能发现的那些问题并没有在这里讨论到。因此,注意查阅本章和其他章节提供的参考书目是十分重要的。本书未曾提到,或只是一带而过的那些内容,在其他参考书中有详细的介绍。

我们以以下四点来结束本章的讨论。第一,在资源许可的时候,应该使用多种试调查法,并进行多轮测试。测试期间发现的每一个不足都是我们在分析和解释时必须面对的一个小问题。第二,学会通过朗读问题和用自己的耳朵来对它们进行评估。我们作为一个研究者,如果对一个问题感到难以卒读、过于矫揉造作或难以理解,那么它肯定会在被调查人那里引起这样或那样的问题。第三,没有任何问题的措辞是神圣不可改变的。如果一个问题在试调查时未获通过,那么我们就必须面对从中发现的问题。然而这时有可能发生两种情况,一种是发现解决这一问题并非一件容易的事,另一种是发现自己是如此钟情于原来的问题形式以致难以割舍。最后一点是查阅其他有经验的研究者编写的问卷实例。但是,在学习问卷设计这门艺术时,我们不可能参考太多的实例。除了教科书和专业杂志之外,许多大的调查组织或联邦政府的统计部门做的调查问卷是可以免费得到的(甚至可以从在线的数据库中得到)。

所有的计算机程序几乎都无一例外地在不断地改进,调查问卷也同样如此。

注 释

1 参见坎福斯和普雷舍(Converse and Presser , 1986)有关插图在调查中的作用,有时这样的调查也叫做析因调查(factorial survey)。在使用这种方法的时候,因子(在例子的括弧中)是变化的。在使用多个因子时,版本数可能会很快增加。为了确定将何种因子用于对这一问题持不同观点的、单独的被调查人群体,我们需要做一些试调查。

设计样本 第7章

Designing the Sample

在抽样调查中我们所做的大多数工作都以常识为依据。例如,在抽样的时候,我们需要考虑我们希望研究的总体是什么,我们可以使用的包括这一总体的清单或资料是什么样的、这些资料的质量如何、我们可能遇到什么样的问题,以及我们怎样才能克服它们等。掌握和了解若干基本原理和用常识进行判断对我们解决上述种种问题是很有帮助的。本章的第一部分主要对抽样问题做比较详细的介绍,说明抽样调查为什么比对每一个调查对象都进行调查更有效,然后对非概率和概率样本之间的区别做了说明。因为概率样本设计是我们的首选,所以本章的其余部分将对它进行重点介绍,内容包括:定义总体、构建和评估抽样框、意外情况及常见情况的处理。与此同时我们也对人们经常提出的"我们需要一个多大的样本"这一问题做了解答。最后我们给大家介绍了几个实例,以帮助大家直观地掌握在计划一个抽样调查时,如何使用普查数据来确定那些重要的、子群体的被调查对象,对于要做的分析其数目究竟多大才是足够精确的。

基本原理

抽样是一种根据既定的法则从一个定义的总体中选取单位元素(elements)的方法。在一般情况下,这些单位元素就是我们研究的对象。它们可以是个人、家庭、耕畜、地块、买卖事务和医院等。我们之所以要抽样,主要原因有两个:第一是将我们感兴趣的问题推广或推论到研究总体(如三月份正在找工作的失业人数是多少?在马里兰州的成年人中,认为在自己的邻里地区,暴力犯罪正在变得越来越严重的人数是多少?),一个实施得当的概率样本使我们得以估计这些数字或百分比。

第二是抽样调查较之普查效率更高,费用更低。普查或试图将总体的所有单位或成员包含在调查中,或试图询问总体的每一单位或每一成员。它需

要 10 年时间进行计划准备，实施 1990 年“人口和住房情况普查(Census of Population and Housing)”和撰写调查报告的费用为 26 亿美元。2000 年人口普查的费用估计超过 50 亿美元。在 1988 年到 1990 年的十多年中，人口普查局提供了 635 000 个临时性工作职位，雇用了 565 000 个人(U. S. Census Bureau, 1989)。据估计，2000 年的人口普查需要雇佣人口普查员将是 1990 年的 2 倍。动用如此众多的人员，尤其是如此众多的、训练有素的人员，其困难程度是可想而知的。然后我们还需要考虑如何对他们进行培训，并建立一整套质量控制检查的规范。虽然人口普查局在这些方面的工作堪称一流，但是这一工作所固有的难以管理操作的这一特点，使得它与一般的大规模抽样调查相比，误差更大、费用更高和能力不足的工作人员更多。实际上，我们曾经在各种场合讨论过用以取代普查的办法，但因这样的更动必须修改宪法的有关条款。而要令国会议员和政府的立法机构相信样本调查具有的种种优越性是很困难的[1]。正因为如此，有关这一问题的争论至今仍在进行。

样本有两种类型:概率样本和非概率样本。概率样本或随机样本是指每一个单位元素都有一个已知的、非零的被选择的机会，且单位元素的选择都是通过某种随机的方法进行的样本。虽然单位元素并不需要有相等的被选择的机会，但是每一个成员不仅必须有相同的机会，而且机会还必须是已知的。在这两个条件都得到满足的前提下，那么再加上正确地使用某些统计公式，我们就可以估计整个总体的各种值和各种估计值的误差幅度。例如，我们可能想要估计校园内，去年冬天至少感冒过 1 次的学生的人数(这一看似简单的研究问题固有潜在的测量问题我们在这里不予讨论，因为在第 4 章我们已经对它进行了讨论)。在我们的研究中，冬天的定义从 11 月 15 日到 3 月 15 日。学校当时在册的学生总共有 28 000 个，我们不仅可以得到全体学生的名册，且名册是完整无缺的。我们准备从中选取 1 000 个学生作为调查对象，最终，全部接到了问卷的学生都按要求填好了问卷(充分的合作)。一个从 28 000 个学生中抽取的随机样本，给每一个学生以 1/28 的被选作研究对象的机会或 0. 035 7 的概率。这两种不同的说法，其含义是完全相同的。

$$\text{选择的概率} = 1\,000/28\,000 = 1/28 = 0.035\,7$$

如果样本中 62% (620)的学生说他们去年冬天曾经感冒过一次，那么我们就可以估计全体学生中，去年冬天得过一次感冒的学生的百分比或者人数。具体的方法是，用选择的概率计算可能的值域的一个估计值，而总体的真值则可能位于这一值域内，这一估计值也可称作抽样误差[2]。为了计算总体中得过一次感冒的人数，我们将选择的概率的倒数(1/0. 035 7)乘以那些曾经得过一次感冒的人数，然后我们再计算这一数字的置信区间。数字变量的置信区间是一个由高于和低于样本估计值的一组值组成的一个值域，该值域中包含了真的总体值。为了计算置信区间，研究者必须首先设定一个概率置信区间，由

其中包含真的总体值的概率,例如90%,95%等。选择的概率0.035 7的倒数是28,这一数字想必大家不会陌生。我们用这一数字乘以得过一次感冒的学生数(620),便可求得总体的估计数17 360 (620×28 = 17 360)。然后,我们必须计算一个置信区间,因为这个样本只是我们从28 000个学生中可以抽取的很多个由1 000个学生组成的样本中的一个。使用95%的概率水平,我们样本的估计值的置信区间为16 518到18 202。这就意味着,如果像我们所举的例子那样,反复地进行1 000个学生的调查,将会有95%的样本的置信区间包含曾经得过一次感冒的学生数的真的总体值。后面我们将会有更多的篇幅讨论总体值的估计问题。而从这一个练习中我们应该了解的主要问题是,样本估计总体值所依据的统计学根据——样本必须是概率样本。

第二种类型的样本是非概率设计的样本。这一类型的样本包括方便(convenience)、立意(purposive)、配额(quota)和滚雪球(snowball)样本。这些样本的名字隐含了研究中的被调查人选取的方法,其具体做法则如下面列举的例子所示。

一个旨在了解哪一个专业的学生——工科的低年级学生,还是社会科学的低年级学生——会在课堂之外学习更多的东西的研究,可以使用方便或立意抽样选取样本。样本中有低年级的工科和社会科学的班各两个。样本的选取不必采用随机方法或令概率为已知的。我们将询问所有班级的所有学生,他们平均每周做多少个小时的家庭作业。因为这些班级都是有目的地进行选择的——在工科和社会科学的低年级的班——所以我们就可以选择那些便于我们安排调查的班,或选择那些指导老师比较熟悉的班,以便比较容易取得他们的合作。在这样的情况下,这两个学科的所有低年级学生就不会有一个已知的被选择的机会,因而也无法估计调查到的结果对所有低年级的工科或社会科学学生的代表性的好坏。

在配额样本中,在调查开始之前,我们首先要确定总体中的哪些特征,如性别、种族、年龄和就业等,可能与我们研究的因变量(dependent variable)有关。在实施一项研究之前,我们希望了解具有这些特征的人在总体中的比例,以便使我们最终的调查能反映总体的分布。如果总体的男女人数各占50%,那么我们希望在最终的调查对象中,男女各一半。如果我们的调查将在某一个特定的市或县进行,那么我们也许可以从美国人口普查或近来进行的一些调查中,得到一些有关总体的信息。对研究对象为学生的研究,信息通常可以从主管注册的工作人员或档案和招生办公室得到。假定我们准备在某一城市做研究,而性别和种族是两个很重要的变量。人口普查数据告诉我们,在18岁及以上的人口中,男性占52%,女性占48%;非裔美国人占27%,白人和其他种族各占68%和5%。假定我们需要做200个面对面的访谈,且总体的种族分布在两性中是相同的。在了解这些百分数之后,我们便可以估计样本的

配额数了：

	访谈数	
	女性	男性
白人	71(0.68×104)	65(0.68×96)
非裔美国人	28(0.27×104)	26(0.27×96)
其他种族	5(0.05×104)	5(0.05×96)
总计	104(0.52×200)	96(0.48×200)

在如上表的配额确定并分配好之后，调查员将被派往指定的地区，如街角、购物中心或选取的街区进行调查。在这些指定的地区，调查员将按分配给自己的配额数完成规定的调查。在调查全部完成的时候，调查样本一般都能在那些重要的变量上与总体完全相配。

这种设计存在的问题主要有三个：第一，对具体的调查对象缺乏控制。我们虽然告诉了调查员被调查对象的类型，但是并没有告诉他们调查对象究竟是哪些人。例如，在一个大型的购物商场做调查时，调查员可能将那些特定的个人，如不讨人喜欢的人、情绪似乎很坏的人或行色匆匆的人都排除在调查之外。这样，那些包含在样本中的人的重要性可能与排除在样本之外的那些人有所不同。第二，我们通常都没有保存接触或拒绝接受调查的人数的记录。唯一保存下来的记录就是完成了调查的人数和我们知道的样本性别和种族特征与总体相配。不仅如此，那些同意接受我们调查的群体也可能与总体中的其余群体有所不同。他们可能对调查持更合作的态度，他们也许更有钱。不仅如此，也许还会存在这样的可能，那就是能否接受调查在很大程度上取决于调查进行的日期和时间，这就可能导致被调查人中的老人或失业的人的比例过大。此外，我们也无法确定，其他那些未曾用于确定目标群体的重要变量，其在样本中的分布是否与总体相同。

第三，同时也是比较关键的问题，这 200 个被调查对象的选择概率是未知的。如果研究是在一个大型商场中进行的话，那么合格人口中那些从来都不去大型商场的人就没有机会被选入样本。相反，那些每星期光顾商场两三次的人就会比那些去商场次数比较少的人有更多的被选入样本的机会。这类批评意见也常见于“街角”或其他类型的以偶然性为根据的样本设计。总之，如果我们希望对总体加以概括，那么总体中的所有单元都必须都有一个已知的、非零的入选我们的研究样本的机会。

另外一种经常使用的非概率设计是滚雪球式抽样。这样的设计大多用于样本单位相当稀缺或难以寻找的时候，其所依据的一个假设是具有相同的特征品质的人可能彼此认识。一个更为普遍的假设是一个不具有某种特征的人也许会认识某些具有这些特征的人。假设我们的兴趣是研究那些 HIV（艾滋病毒）阳性的人，了解他们的生活方式在哪些方面与常人不同，以防止人们受

到艾滋病毒的攻击。不言而喻,这一群体是一个稀缺而难以寻找的群体。这样一个群体的抽样框是不存在的,尽管这一群体的人数正在上升,但他们仍然是总人口中相当小的一部分。为了进行这样一个研究,我们可能将两种非概率抽样设计合并使用。第一,我们需要确定或得到进入可以找到 HIV 阳性的人的组织或团体的途径——例如,乐意的男人们或 AIDS 援助团体的组织。依据研究的日程和已有的资源,我们也许可以得到一个方便或立意的样本。通常我们可以借助在公告板上贴通知,或在一次公开的集会上陈述调查的原因,得到与某一组织的部分或全体成员进行接触的允许。尽管自愿接受我们的要求的人可能只是少数,但是这毕竟给了研究者一个出发点。随后的一个步骤便是进行调查。在调查结束的时候,调查员可以采用"雪球"策略,要求被调查人提供他所认识的满足我们条件的人的姓名、电话号码或住址,以便接着进行调查,这就是说,要求一个总体中已经确定的成员确定总体中其他的合格人选。

而滚雪球设计隐含的前提——具有类似特征的人有着千丝万缕的联系是确实无疑的——正是一个非概率样本的通病。最终的样本究竟对 HIV 阳性人口的总体代表性如何?我们无法回答这一问题,因为我们不知道选择的概率,因而无法估计抽样误差的边界或限度。此外,我们不知道如何限定初选的组织或志愿者。我们既不知道与我们配合的组织或志愿者是否是研究范围内所有的 HIV 阳性的人?也不知道组织的成员或志愿者的交往是否是仅限于某些封闭的小圈子?我们无法估计样本结果的误差范围,因而会严重损害最终研究结果的可信度。

然而,非概率样本对于某些目的的研究还是非常有用的。并非每一个研究都需要对总体的某些特征进行估计,或加以概括。在探索性研究中,研究者也许只是希望了解人们对某些问题的想法、信仰程度或感觉,这样的信息可能对以后涉及一个规模较大和更为全面的研究很有用。非概率样本在问卷题项的早期测试中也很有用。我们可能无法确定某些特殊类型的人是怎样理解某些问题的文字含义的,这样我们不妨将我们的试调查只是集中在这类人身上就可以了。或者,假如我们事先对总体或兴趣群体的情况有所了解,那么我们就可以利用这些信息,以便更好、更集中地使用我们手中的研究资源。例如,如果我们正在为医务工作者设计一种教育模式,其目的主要是纠正他们对艾滋病的某些最严重的误解,那么我们就可以去调查那些对艾滋病略知一二的群体,而非对所有的医务工作者进行调查。这里,我们所持的假设是,一个对艾滋病有误解的群体,如曾经看护过艾滋病受害者的护士群体,也可作为其他医务工作者群体的代表。因此,艾滋病护士的研究结果将可以被推广到所有的医务工作者中。

定义总体

让我们回到概率样本的讨论。在建立概率样本的过程中，两个最为重要的问题是定义总体和挑选或构建抽样框。所谓总体是指我们希望研究的单元的聚合或组合。我们研究的结果将被推广到这一聚合或组合。我们从研究的问题开始我们的概率样本的讨论。我们希望研究的问题是什么？例如，我们可能希望了解"成年人中认为不管出于什么原因，妇女都应该可以做流产的人的比例是多少？"

在定义总体的时候，分析的单位或单元和单位边界的定义这两个问题是必须明确的。就分析的单位而言，我们必须确定研究的问题所关注的是个人住户、家庭住户、集体住户(group quarters)还是其他类型的居住单位。个体一词是不言自明、无需定义的，而其他两个用词则是必须加以说明的。根据美国人口普查局(the U・S・Census Bareau,1993)的定义：

> 一个家庭住户包括所有居住在同一居住单位中的人。一个居住单位是一所房子、活动房屋、一组房间或一个单独的房间，是人们单独的居住场所。单独的居住场所是居住者的起居、生活和进餐与他人分离，且通过一个公共的厅堂，直接进出建筑物的那样的居住单位。居住者可能是一个单独的家庭、独自居住的个人、在一起生活的两三个家庭，或其他彼此相关或不相关的人，由共同协议安排居住在一起组成的住户。(p. B-13)

集体住户是各种类型的集体安排的住户。美国人口普查局定义了两种类型的集体住户：公共机构的和非公共机构的。公共机构的集体住户是居住者都受到正式授权的护理或监管式的监护或监控的场所。如国家监狱、联邦感化中心、军事禁区、地方监狱、警察的拘留所、途中的小客栈、疗养院、精神病院、弱智学校或病房、遗弃儿童之家和青少年犯罪管教所等。美国人口普查局对非公共机构的集体住户的定义是 10 个或 10 个以上相互没有关系的人安排居住在一起的住户。这种形式的住户有公寓、提供照料和服务设施的集体户，如未婚母亲的孕妇之家、宗教集体户、海船上的海员或单位的员工宿舍等。调查人必须在研究的设计阶段确定什么样的住宅居住安排形式含有符合自己研究所需要的被调查对象。

我们怎样确定自己的研究单位是个人、家庭户，还是其他什么单位呢？这主要取决于我们的研究目的。如果我们研究的问题是对流产问题所持的态度，那么个人也许是比较恰当的分析单位。因为态度可能因性别或年龄而异。因而将家庭户设计为分析单位可能是不恰当的。如果态度在家庭户内的不同成员，如夫妇或家长和子女之间存在差异，那么我们就不能轻易地报告家庭户

的态度。虽然态度在本质上是个人的，但去年的总收入或诸如冰箱和小轿车这样的耐用消费品的拥有或购买数这样的问题，最好以家庭户为单位进行研究。虽然个人和家庭是两个常用的分析单位，但组织或机构也有可能成为我们的分析单位，例如工会组织、学院、环保团体、医院和俱乐部等都有可能成为我们的分析单位。一项旨在研究小学科学设备种类和质量的项目，适当的分析单位当然应当是小学。一个估计过去5年中器官移植数和种类的研究，其分析单位当然应当是医院。分析单位的确定应该取决于研究者，但这一决定必须在研究的设计阶段做出。

定义总体要做的第二项工作是设定定义的边界。对于一个个人的总体而言，设定通常包括以下几个方面：研究地区的地理边界、研究总体的地理特征、调查对象的年龄及包括的居住单位等。研究地区的地理边界是由研究将要概括的总体和可资利用的时间及资源决定的。依据这些限制条件，研究地区可以定义为一个城市、县、大都市的统计区域，或某些其他的地理实体。地理区域必须在设计阶段的初期决定，因为它还关系到其他许多问题，如得到一个抽样框、招聘面对面调查的调查员和估计电话调查的电话费等。

我们也必须确定合格的被调查人的特征和他们可能的居住单位。例如，全美综合性社会调查（the General Social Survey，GSS）是一项全国性的概率抽样调查。它的调查对象是美国50个州的家庭户中居住的、操英语的成年人。有关这一定义有几点必须加以说明。第一，这一研究包含了美国全部50个州。许多面对面调查都将阿拉斯加和夏威夷排除在外，因为在这两个州，培训、督导和维持调查员的费用比较高。这两个州的人口只占美国人口的0.065%（不足1%）（U.S. Census Bureau, 2000d），而许多调查者愿意采用更为严格的总体定义。第二，这一调查只包括了操英语的成年人。在1987年到1991年间，将会排除将近2.5%样本家庭户（Davis & Smith, 1993）。这些被排除在外的家庭户代表了操13种语言的不同群体，其中60%都操西班牙语。为了将这些群体包含到我们的调查中来，我们需要做不少的事情，而为了做这些事情，每一个调查所需的费用将明显上升。为了使我们的调查能包括操非英语的被调查人，我们首先必须了解有操什么样的语言或方言的群体出现在了我们的样本中。德语、汉语、土耳其语、阿拉伯语、越南语或其他语言群体是否会出现在我们的样本中？在掌握了语言群体的种类之后，我们需要为每一种语言做问卷的双向翻译。先将问卷从英文译成外文，然后再将外文文本译回英文，以确保译文正确无误。此外，不仅需要编写和/或修改调查员培训手册，而且还需要聘用、培训双语调查员，并将他们派到非英语住户那里做调查。不言而喻，做这些事情不仅仅使调查费用有所增加，而且也许还会延长调查进行的时间。

第三，必须考虑合格的被调查对象的年龄。成人可以定义为任何18岁及

以上、19 岁及以上,或 21 岁及以上的人。究竟定义为哪一个年龄段,完全取决于我们的研究目的和我们希望对之进行概括的群体。有时,只有那些处在某一特定年龄段的群体的人才是合格的调查对象,如所有年龄在 18 ~ 64 岁之间的人,或年龄在 21 ~ 49 岁的妇女等。前一个年龄群体,我们的兴趣可能在那些尚未退休的人,而后一个年龄群体,我们的目标则可能是了解妇女在高中毕业之后,是否参加工作。如果参加工作,那么在什么年龄,从事何种职业。

合格的被调查对象居住的地方的居住类型是我们应当考虑的第四个问题。GSS 调查的是家庭户,这就意味着那些居住在宿舍、收容所、疗养院和其他类型的集体住户单位的人将被排除在调查之外。在 2000 年,美国人口中的 2.76% 居住在集体居住单位中。但不同年龄组居住在这种类型的居住单位的人的比例有明显不同(U. S. Census Bureau, 2000d)。例如,在 1980 年,年龄在 18 ~ 24 岁和年龄在 75 岁以上人口中,居住在集体居住单位的人的比例分别是 9.4% 和 11.4%,后者比前者高 2 个百分点,但后者比 25 ~ 64 岁年龄组的这一比例只高了 1 个百分点稍微多一点(Davis & Smith, 1993)。2000 年,18 岁以下的青年,居住在集体单位的比例为 0.4%,而在 18 ~ 64 岁和 65 岁及以上的人口中,这一比例分别为 3.1% 和 5.7%。

总体究竟应该多大取决于若干种因素。为了确定总体的范围,我们必须回答这样一些问题:哪些群体是我们希望加以概括的?其他研究者在研究这一问题时,是如何定义这一总体的?我们能用于这一调查的资源是什么(资源不仅包括时间和预算,也包括人员和设备)?是否存在进行这一研究或定义总体的"最佳方法"?是否可以取得含有定义总体的合格成员的清单或信息源?

构建抽样框

概率抽样的总体一经定义,接下来马上要做的工作便是寻找或建立抽样框。这种框是一个(或一些)含有定义的总体单元的清单或信息源。在电话调查中,框可能是一本电话簿,或一份含有该地理区域的所有电话前置号码的清单。在邮寄式问卷调查中,框通常是一份登录有所有单元的清单,如组织成员清单。如果研究的目的是确定近年来毕业的学生对他们在学院或大学受到的学术训练的满意程度,那么我们也许可以从校友办公室或注册办公室得到一张近 5 年来毕业学生的名单。读者是否能设想一下,在使用这些类型的清单时,可能会遇到一些什么样的问题(我们将在后面讨论这一问题)?

一种经常使用的抽样框是人口普查数据汇编。在没有合适的、完整的清单可供我们定义总体使用的时候,我们可以使用这些数据。人口普查数据通

常用来作为一般人口总体，例如18岁及以上人口的研究的面对面调查的抽样框。您可以认为美国所有成年人的清单的确存在。然而，实际上只有两种比较完整的清单，且这两种都是保密的，并不对公众开放：10年编辑整理一次的人口普查表和每年填写一次的所得税表。其他可能可以得到的清单都有重大的缺失。例如，电话簿不包括那些没有电话，或虽然有电话但没有登录的住户。来自州驾驶执照管理局的清单不包括那些不驾驶汽车的人，选民注册清单排除了1/4～1/3的成人，而财产税记录或动力公司的文件不仅没有包括全部成年人，且可能也不能作为公共信息使用。

在姓名和住址保密的情况下，我们怎样用人口普查的数据选取成人的样本呢？这的确是一个值得我们认真思考的问题。一个不使用复杂的抽样理论的简单明了的例子将会非常有用。（因为这一过程的部分内容所描述涉及的抽样理论，超出了本书介绍的范围，所以你们只能根据对作者的信任，来接受本书介绍的某些有关结果。）

我们从研究的问题着手来讨论这一问题。假定我们想要研究成年人对流产问题所持有的态度。那么我们随后应该怎么做呢？我们需要对成年人的含义加以定义，并设定定义的边界。我们假定我们感兴趣的调查对象是年龄在18岁及以上的居住在家庭户中，且我们有足够的资源实施一项全州范围的面对面调查。我们再进一步假定，我们要求的样本量是1 000个完成的调查，且每一城市街区平均完成的调查数或最后一阶抽样完成的样本量是5个[3]。

我们使用人口普查的数据分阶段抽取样本。在第一阶段，我们选取县或县的群体。在这一项目中，我们想抽取24个县。这24个县的抽取，采用概率比例抽样，比例以每县的18岁及以上的人口数来定[4]。在每一个县内，城区抽取城市的街区，农村则抽取地块。这一阶段的抽样单位的抽取也是以成人人口数为比例的概率抽样。

注意，我们已经进行了两个阶段的抽样，其根据仅仅是没有姓名和地址的人口普查的数据。然而，在最后一阶，我们必须计数在第二阶段抽出来的街区或地块的住户。我们将派制表人(lister)前往每一个街区或地区，编制每一个住户的地址或所在位置的清单。在编制完成之后，我们将对制表人发现的街区或地区的住户数，与用于街区或地区抽取的那一数据（人口普查数据）进行比较。如果我们希望回答率为70%，且假如预先确定的每一街区应完成的访谈数为5，那么每一个街区大约应该选取7(5/0.70 = 7.14)个调查对象。采用被定义的地理区域内的人口或住户的计数，并在最后一阶段抽样时由制表人到实地记数，是在开始抽样之前我们手中并不掌握总体清单，但又必须使用概率抽样时使用的一种切实可行的方法。最后一个应该注意的问题是：每一阶段使用的数字都要足够精确，或不会导致未知的偏倚。

定义的总体与抽样框的匹配问题

在很多情况下我们无法找到与定义的总体完全匹配的抽样框，这样我们必须采取折中的办法。一个成年人的电话调查便是这一问题的明证。因为并非所有的成年人都有一部电话，且即使有电话，也未必全部都登录在电话簿上，因而我们无法找到一个完备的抽样框，因为它并不存在。现在我们再回到有关成年人对流产问题所持态度的研究，以具体地阐明这一问题。

假定准备在北卡罗莱那州的威克县做这个调查，该县辖有罗利市。我们决定采用电话调查进行我们的研究。我们应该采用什么样的抽样框呢？它可以使该县每一个成年人以一个已知的概率入选我们的样本。我们首先想到的是电话簿上的白页，其实这的确是一个不错的主意。然而，作为一个抽样框，它存在种种不足。首先它没有包括那些没有电话的人。但是，我们已经决定采用电话调查，我们就不得不接受这一可能的偏倚。在北卡罗莱那州的威克县，大约有 1.2% 的住户家中没有电话。问题的关键在于"那些住在没有电话的住户中的人，是否与那些被调查的有电话的人有什么不同？"一般讲，差别肯定是有的。然而，他们在总人数中只占了相当小的比例，而我们设法将他们包含在我们的研究中所需要的费用（通过面对面访谈），将大大超过他们对最后的结果产生的价值，既然如此，我们有什么必要一定要将他们包含到我们的调查中来呢？所以，我们将容忍这一可能的偏倚。

第二，电话簿并没有包括那些未将自己电话号码登录在电话簿上的人。在罗利—达拉谟—卡里联合统计区（Raleigh-Durham-Cary Combined Statistical Area），包含了威克和其他七个县，没有电话的住户群体大约占全部住户的 22% 。不言而喻，我们不能对这么大的一个数字视而不见。然而，我们将要在第 8 章介绍的随机号码拨号（random-digit dialing，RDD）抽样法，能使这一问题迎刃而解。用电话簿做抽样框，或一般的电话调查存在的第三个问题是，有些住户家中拥有一部以上的电话，他们会因此而具有更高的被选择的概率。这一问题是可以弥补的，只要我们在调查时加上一个问题，问一下住户家中有多少部电话，将营业性的号码去掉。来自那些有一部以上住宅电话的住户的数据可以电话数目的倒数加权，这样我们在分析这些数据时，每个住户便有了相等入选我们的调查的概率[5]。

第四个问题与罗利市的电话簿覆盖的地理区域有关。电话簿涵盖了该县的多大地理区域？全部、大多数、一半，还是超出了威克县的范围？这一点是需要加以确定的。在审视了电话簿之后，我们得知，电话簿上的号码可分为两个住宅段。一段包括罗利市及一些城镇周边地区和一些行政区划未定的地

区。这一段大约覆盖了该县地理面积的68%，由一家电话公司提供服务。第二段包含的地区，则由两家以上其他的电话公司提供服务。这一段包括了该县的其余地区、一些小城市（人口不足5 000）和县外的农村地区。

我们共有三种选择。一种是只选取电话簿中的第一段，它包括罗利市和大约68%的县的人口。我们可以重新将我们的总体定义为罗利市及周边地区的成年人。在给我们的研究做结论的时候，我们必须要对这一总体的特征加以描述。我们可能需要说明生活在城市或城区的成人的比例、农村中成人的比例、这一总体收入的中位数、平均受教育年限等。

第二种选择是同时使用电话簿上这两个部分的号码。为了使抽样框和我们的总体定义相匹配，我们必须对某些电话号码进行筛选。尽管电话簿上已经列出了该县所有的前置号码，并指明了这些前置号码所特有的每一城市，但是它却未曾指出那些使用范围超出城市或城镇边界的前置号码。因此，对于所有的调查都必须是在城市边界内的所有城镇中进行的调查，我们都必须询问每一个被调查人，他们的住处究竟位于威克县还是相邻的县。从人口普查的资料我们可知，大约8%的样本号码必须以这种方式进行筛选。如果被调查人的住处不在威克县内，那么他就不是合格的调查对象，我们就不应该继续对他进行调查。

第三种选择是将电话簿上的两个住宅段都定义为我们的总体。在我们再一次查阅人口普查的数据之后，我们估计电话簿上的号码大约有4%左右来自邻县富兰克林、约翰斯顿和哈涅特。最简单和最完整的方法是将总体定义为登录在威克县电话簿上的住户。这将包括所有威克县的住户和县界之外的少数几个小镇。这一定义将使总体的数量增加4%。因为威克县对于我们研究的问题并没有什么独特之处，所以我们没有必要使我们的样本框与县的边界完全匹配。然而重要的问题是，在做研究报告的时候，我们必须给读者提供我们的研究总体的确切和精确的定义，并对这一总体的地理特征加以描述。

认识抽样框中存在的问题

我们从多年的调查中得到的经验是决不要轻信一个抽样框。因为不完整和不精确是一个抽样框所固有的弊病，所以研究者必须认识到它可能存在的问题，并懂得在它们发生的时候，如何进行处理。在框是一份清单的时候，通常有可能存在以下问题：

- 清单中含有不属于总体的单位——这样的单位称为不合格单位。
- 个体或单位的信息不够精确。
- 个体或单位的信息丢失。

- 某些个体或单位重复登录。

不合格个体或单位

一种常见的问题是不合格个人或单位。在遇有这样的情况时，我们怎么办呢？答案很简单，那就是不对他们进行调查，但把他们包含在抽样率中。抽样率等于为了达到要求的样本量，我们必须从抽样框中选取的数除以在抽样框中总的抽取数。用一个例子来说明这一点可能比较容易理解。假定我们希望对我们的大学近年来的毕业生做一个他们对于5年前开始的课程改革的满意度调查。我们只想对最近3年毕业的学生进行调查，但我们唯一可得到的抽样框，却包含了过去6年毕业的毕业生。这时，我们应该怎么做呢？我们可以使用这6年的框，但是我们应该确定最近3年毕业的学生占的比例是多少。为了使事情简单易行，我们假定近3年和后3年毕业的学生相等，各占50%。如果清单的排列不是按毕业年份排列的——整个清单中，两个部分的学生的排列是随机的——那么我们可以预期一个包含1 000个姓名的随机样本，两个时期的人数应各约为500。因为我们对毕业时间在3年以上的人不感兴趣，因此每当我们选到这部分人的时候，我们只需将他们忽略不计，继续进行下一个抽取。尽管如此，我们还是需要对清单中不合格人选的数目做一个估计，以便计算抽样率。清单上存在不合格人选，并将他们忽略不计，并不影响选择的概率。其理由很简单。如果清单上有过去6年间毕业的学生18 000个，每3年毕业的学生各9 000。那么抽样率便是1 000/18 000（18取1）或0.055。假如我们从总共有9 000个合格人选的清单中抽取500个人的样本，抽取概率为500/9 000 =0.055。可见，在目前这一情况下，不合格人选虽然是抽样率中的一个部分，但他们并未对选择的概率发生影响。

然而，我们常常遇到的情况却是不合格人选的确会对选择的概率产生影响。而这样的问题之所以产生，大多是因为抽样方法不当。假定合格人选和不合格人选混在一起排列在一张清单上。研究者常犯的错误是忽视不合格的人选，并直接抽取下一个合格人选，而不是根据清单上的不合格人选的比例修正抽样率。你们是否能指出这种做法究竟错在什么地方？这种做法的错误在于，给了那些在清单上先于一个不合格人选的合格人选以双倍的入选机会：即他们自己入选的概率和不合格人选的入选概率。

这种错误的做法常见于在用电话簿进行抽样的时候。尽管在用电话簿进行抽样的时候，我们的目的是抽取一个电话号码的样本，但是通常抽取的是刊登的行和电话号码的样本。因为许多电话簿将住宅的条目和营业场所的混在一起，为了选取住户的样本，我们必须将营业场所的条目作为不合格电话处理。另外一个经常发生的问题是，名字或地址比较长的那些人，在电话簿上被登录为两行，致使电话簿上的每一行未必总是一个电话号码（见第8章表8.2）。

没有人会先计数电话簿上的合格的电话数，然后再从登录的合格号码中选取一个简单或系统的随机样本。我们不妨来看一下如芝加哥那样的大城市的电话簿。它登录的条目共有 1 400 页，每页为 5 列，每列有 110 多行。要计数一下电话簿上登录的所有的人，可能需要一个星期。这时研究人员需要做的不是计数全部登录的人，而是先估计那些不含合格的住户号码的行的比例，再将这一比例计进抽样率，然后再从电话簿上选取行的样本。尽管这种方法是正确的，但是我们仍然必须小心从事，给每一条条目以相等的选择机会。这样，假如营业场所的条目和住宅电话的混在了一起，我们就需要估计一下营业场所的条目的比例。对于空白行和其他不合格的行，我们也必须做同样的处理。总之，我们必须在求抽样率之前，得到不合格行和空白行的总数的估计数。这一估计数将是抽样率的一部分[6]。问题的关键在于确保抽样单元的选取概率不会因为选择方法的变化而变化。

不精确性

抽样框中常见的第二种问题是有些信息不够精确。也就是说，有的时候我们也许自以为选取的是一个合格的单元，但后来证明它是不合格的。因为所有个案的抽取都是完全在一起进行的，我们事先并不知道哪一个是合格的，这样，不合格的单元便变成了一个问题，这不仅是因为每抽一个个案都需要花费一定的时间和费用，而且它还会减少我们在计划的时间和预算内得到的合格的个案数。因此，我们必须尽早发现这一类型的问题，最好在试调查之前或之中发现它们，这样我们便会有机会采取措施予以纠正。

抽样框可能存在的另一个问题是，在一个住户或一个电话号码内，可以找到的合格调查对象不止一个，而是几个。假定我们登录了某一城市街区的所有住户，并从中选取了将要对之进行调查的住户。调查员在到达某一地址时，发现那里有两个住户，而非一户，前面一户是已经登录的单元住宅，而后面是一户没有登录的单元住宅。这时调查员应该怎么做呢？在这种情况下，正确的做法是对这两户都进行调查，因为第二户也是合格的，尽管它没有登录。我们希望所有的合格单位都有一个被选入样本的机会，而这是第二户唯一被选入样本的机会。此外，我们希望对所有的合格单位加以概括，而这第二个单位也是一个合格的单位。在多数情况下，我们一般可以假定这样的错误是随机的。因为这是一个我们偶然才发现的（随机的）错误，因此其他类似错误也可能偶然被发现，因而也是随机的。这也可能如上面所阐述的那样，有些以为是合格的单元，实际却是不合格的。如果框是最近的，且质量很好，那么这样的情况就可能不会经常发生。不过，掌握恰当的纠正方法仍然是十分必要的。

丢失信息

那些在我们意料之外的单元可能会成为一个比较大的问题。在作者本人

于 1987—1988 年间进行的一项研究中,最容易得到的抽样框是人口普查数据。我们决定使用这一数据,但是在那些高速发展的地区必须对它进行必要的更新,因为这个数据已经过去了 7 年。我们用作更新的资料来源主要有:市政当局编撰的有关新住宅区和拆迁的老住宅区的统计资料、我们与银行、房地产公司联系得到的有关该地区发展情况的资料等。以往的经验告诉我们,没有任何一项资料包含了我们所需要的全部信息,最好的办法是使用多种资料。使用这些来源不同的资料,我们对 1980 年人口普查的数据进行了更新,然后再以这些更新的数据为依据,抽取了街区。除了一个街区之外,几乎所有抽出的街区的情况都很好。我们本来估计那一街区可能有 50 个住户,但是在登录员到达那里的时候,却发现那里有一座高楼拔地而起,并已经为住户所占。也就是说,在调查时,那里已不再是 50 户人家,而是有了 450 户人家。在前面那一个例子中,原来我们预期有一个住户,但实际发现了两个,我们对两户都进行了调查。而在这一情况下,如果我们也采用类似的办法,这就意味着对这 450 户采用我们已经用于 50 户的那个抽样率。如果我们使用每 7 户抽一户的比率,而不是抽取 7 户,那么我们大约总共要抽 64 户。如果合作率为 70%,那么我们将在这一街区实际完成 45 个访谈。因为我们计划调查的总数是 500,那么这一街区最终完成的调查数,将与我们完成的总的调查数不成比例。在遇到类似这样的情况时,我们可以允许这一街区做两倍或三倍于其他街区的访谈。例如,我们可以允许在这一街区做 15 个访谈,那么我们可以用因子 3(45/15 = 3)给这些访谈加权。

在被调查人选好之后,也可能会发生一些我们意想不到的事情。在最为一般的人口总体中的个体调查中,每户只选一个被调查人做调查。在一项以个人为单位的研究中,我们是怎么来考虑这一问题的呢?是否所有的合格人选都有相等的入选概率呢?答案是否定的。在一人户中,户中唯一的人自然而然地成为了我们的调查对象。但在有三个合格人选的户中,只有一个人被选为调查对象,那么在这一阶段,每一个人入选的概率就是三分之一。可见,入选的概率是随合格人数的变化而变化。为了使概率对所有的个体都相等,数据必须加权,以反映入选住户中合格的人数。在一人户中,权数是 1,两人户的权数是 2,三人户的权数是 3,其余户型,依此类推。

重复登录

抽样框另一个常见的问题是某些个体登录了一次以上,除非我们采取某些纠正措施,否则这些个体选择概率将高于其他个体。这样的情况之所以发生在电话调查中,是因为某些住户,由于各种原因,拥有一个以上的电话号码。可能是:一个号码给十几岁的孩子,另一个号码则是家中的成年人的;住户可能有第二个电话号码,专门留给关系比较密切的朋友或亲戚;或者其中一个号

码是用于营业的。联邦通信委员会(The Federal Communications Commission)估计,约6% ~8%的美国家庭拥有一部以上的住宅电话[7]。纠正这一问题的办法是,在调查中增加一个有关住户中有几个不同的电话号码的问题,并以户中拥有的非营业性的电话号码数的倒数为权,对被调查人的数据进行加权。如果户中有两个非营业性的电话号码,那么数据的权数为0.5,有三个则权数为0.33,其余依此类推。

类似的情况在使用大学师生名录时也可能发生。有些学生或教职员可能登录在一个以上的部分,那些享受助教奖学金的研究生,既登录在学生名册部分,也登录在教职员部分。每当我们抽取了一个学生或教职员时,我们都必须看看名册中的其他部分,弄清楚这个人是否被二次登录。我们可以给所有登录了两次的那些人用0.5加权,或随机地删除这些个案的一半。这些方法将使选择的概率保持相等。

最后一个人们经常会提出的问题是:“样本框究竟应该需要有多精确和多完整?”对于这一问题,我们无法给出确定的答案。就不精确性而言,主要的问题是我们需要一定的时间和费用,对每一个个案加以甄别,以找出不合格的被调查人,或也许是找到一个合格的被调查人,但已经搬到调查的目标地理区域另外一个地点居住。这样做一般都需要一点额外的收集最近的消息和安排进行调查的时间。而诸如这样的尝试都可能消耗一些用于调查合格的个案的资源。正因为如此,花费在不精确信息上的时间和精力一般都会减少那些本来有可能进行调查的合格的个案数。

抽样框究竟应该有多么完整?这一问题同样也很难回答。研究者必须考虑并回答的问题有若干个。丢失的单元与包含在抽样框内的单元的差别究竟有多大?丢失的单元对最终结果的影响究竟有多大?它只是对单个变量的结果有所影响,还是对变量间的关系也发生影响?在应用一个抽样框之前,我们必须对它的不精确性进行必要的调查,并确定它是否会对我们的结果产生影响。

确定样本大小

在抽样调查研究中人们问得最多的问题之一是:“样本究竟应该有多大?”与许多类似的问题一样,这一问题不存在简单的答案。样本的容量是若干个因素的函数。它们包括采用的研究设计、关键变量的变异性、是否要估计总体值、是否要检验假设、两个变量之差和它们的差的标准误差的大小等。在进行假设检验的时候,研究者在确定样本大小时,力图使用数据所做结论的两种类型的错误达到最小。这两种错误是:①在两个变量实际无关时,声称它们

相关(称为甲类错误);②与第一类错误相反,在两个变量实际有关的时候,声称它们无关(称为乙类错误)。解决这些问题和确定样本量是相当复杂的,科恩(Cohen,1988)、福莱斯(Fleiss,1981)及摩斯和卡尔登(Moser & Kalton,1972)等人曾对这些问题进行过深入的探讨。在开始探讨怎样确定样本量之前,我们先使大家对它所需要的基本信息有一个大致的了解。为此,我们想先用一些例子来阐述在估计总体百分比时使用的确定样本量的经典方法。

确定表示为百分比的变量的样本量的近似公式(Cochran, 1977)是:

$$n = \left(1 - \frac{n}{N}\right) \times \frac{t^2(p \times q)}{d^2} = \text{有限总体纠正因子} \times \frac{\text{概率水平} \times \text{方差}}{\text{置信区间}}$$

公式中

n = 样本量或完成的合格单元的调查数

N = 合格总体的大小

t^2 = 与值的正态分布下的面积有关的标准差得分的平方值

p = 用于计算样本量的类别的百分比

$q = 1 - p$

d^2 = 围绕样本估计值的一半的精度区间的平方值

这一公式可用语言表述为:样本量等于有限总体的纠正因子($1 - n/N$)乘以这一样本发生的概率水平,乘以方差或总体中的变量的变异性,然后除以我们希望的估计值的置信区间的大小。我们先来更深入地看一下公式的这些组成部分,然后再来看一个实例。

有限总体纠正因子

有限总体纠正因子(finite population correction,fpc)是一种修正因子。在采用无回置法选取样本单元时,它已经成为公式的一部分[8]。然而在样本量小于总体总数 5% 的时候,fpc 对最终结果的影响非常小。在大多数样本调查中,一般都不考虑 fpc。在下面的例子中,fpc 就未曾予以考虑。在较后的例子中我们还将更多地讨论样本量和总体容量的关系问题。

概率水平

这时公式还有三个组成部分。为了求方程的解,我们设定其中两个组成部分(概率水平和置信区间)的值,并大约估算一下第三个部分(方差)的值。t 值,概率水平是标准差的得分,它表示在变量呈正态分布时,变量的值落入一组区间的百分比。1 个标准差大约包括了 68% 的样本值,而它的得分是 1.0;2个标准差包含了近 95% 的样本值,而它的得分是 1.96;3 个标准差则包含了大约 99% 的样本值,它的得分是 2.58。在使用这些值或得分时,我们假定我们的变量是正态分布的。设定 t,就意味着我们在为我们的样本结果设定

一个概率水平。这就是说，在我们将要进行的100个以上这样的样本量的调查中，我们希望有百分之x(x是我们设定的)的把握确信，我们的样本的置信区间(d)包含了总体值。

方差

这一公式将一个变量表示为一个百分比。一般讲，之所以这样做是因为我们对一个表示为两种类别的变量感兴趣，这两类可以是是与不是，或有与没有等。例如，我们也许先估计总体中抽烟、拥有枪支、乘坐公共交通工具，或年收入在3万美金以上等类型的人数的百分比。这些变量中的每一个都以是或有和不是或没有的百分比来表示的。我们也许对以非常同意、同意、不同意和坚决不同意表示的答案类别的变量感兴趣。在决定样本量时，我们可以通过确定哪一类或哪些类是我们所我们感兴趣的类别，将这一变量二分化。如果我们对同意一种观点的人数的比例感兴趣，我们可以把非常同意和同意两类答案合并成一类，而把两个不同意的类别合并成另一类。最终得到的两个类别的百分比之和必须等于1.0。一个百分比变量的方差等于这两个百分比的积。

如果我们希望做一个研究来估计罗利的成年烟民的人数，为了确定样本的容量，我们必须在研究开始之前，大致猜测或估计一下烟民的百分比[9]。这一估计值是计算样本量的公式不可或缺的组成部分。虽然这样做看起来与我们承担的研究任务似乎不太相宜，但是我们别无选择。可用于这样的估计方法有若干种。一种是用疫情研究中得到的结果。如果以前的调查地点的特点与罗利不甚相符，我们可对以前的估计数做一些修正。另一种方法是对烟民的比例做一个保守的估计。诚如我们很快就要向读者阐述的那样，方差部分对公式的解的影响并不比置信区间大。因此不够正确的估计值对调查需要的样本量的影响，并不如人为设置的那个围绕在估计值两边的置信区间的精确度那么大。第三种估计方差的方法是做一个小规模的探索性调查。根据我们手中掌握资源多少的不同，对一个样本量在25~50的住户的随机样本进行一个简短电话调查，设法得到一个合乎情理的烟民比例的估计值。在这一例子中，p表示抽烟人数的百分比，而q则是$1-p$。

置信区间

公式的第三部分是置信区间，即d，它由研究者自行确定。这一区间是我们要求或容忍的误差幅度。从理论上讲，无抽样误差地估计总体值的唯一方法是包含或调查总体的每一单元。在我们调查一个个案的样本时，我们只可以估计总体值位于某一个值域之内。d以正负半个值域表示。因此，如果假定罗利抽烟成人约占人口的30%，那么d就是我们需要的值域在30%左右两

边的大小。正如我们很快就要阐述的那样，这一值域越小，需要的样本就越大。

让我们先来回顾一下估计样本量的公式的所有组成部分，然后再来看一个实例。方差是一个我们必须猜测或估计的值。具有某一特征的总体单元的百分比是 p，而其余的单元的百分比是 q。概率水平 t 是我们选定的，表示在 100 个反复抽取的样本量与我们的样本完全相同的样本中，总体值可能位于特定的置信区间内的样本数有多少。置信区间 d 是我们希望的我们的估计值落入区间的大小。

把各个部分合在一起

让我们来完成估计罗利市成年人抽烟人数这一实例。我们的问题是"如果我们假定公式的组成部分的构成如下，那么我们需要的样本量是多少?"我们假定罗利抽烟的成人的比例与全美人口中这一比例相似。我们假定的理由是，罗利既是一所著名大学的所在地，人口教育水平高于全国的平均水平，也是一个烟草主要种植州的首府。我们假定这些因素彼此抵消，因而有理由认为全美抽烟的成年人的百分比是罗利这一数字的一个不错的估计值。我们假定 $p = 0.30$，因此 q 就是 $1 - 0.30 = 0.70$。我们希望我们的概率或置信水平等于 95%。即每一组样本量相同的 100 个样本，能有 95 个样本包括总体值。同时，我们希望我们的置信区间为 ±5%。总之，问题概括起来可表述为这样："如果我们有 95% 的把握确信总体的值位于一个 25% ~35% 的区间内，那么需要的样本量是多大?"将这些值代到公式中，我们得到：

$$n = \frac{(1.96)^2(0.30)(0.70)}{(0.05)^2} = \frac{0.8067}{0.0025} = 323$$

我们需要调查的数目是 323。记住我们的方差估计值是一个猜测值或大致的估计值。如果我们的确做了 323 个访谈，但却发现我们的估计值是错误的，那么它将对我们的计算有什么影响呢？答案是置信区间将会受到影响。做此类计算的目的是获得一些有关必要的样本量的认识，即在基于我们研究的问题而建立的限制条件内，我们需要的样本究竟应该有多大。但是我们必须了解，诸如这样的计算只不过是引导性的。我们并不需要真的刚好去做 323 个访谈。然而，我们的方差估计值如果正确的话，那么为了达到规定的精确度，我们就应当争取完成与这一数目相当的访谈数。

调查一旦实施之后，我们总是要为关键的变量计算置信区间。置信区间与每一变量的变异度、完成的访谈数和我们设定的概率水平有关。在上面的例子中，如果总体抽烟人数的比例的估计值过高，就意味着方差较小，我们最终的置信区间将会小于 ±5%。这将会使我们的样本估计值的精度高于我们所要求的。如果我们的估计值过低。那么就意味着实际的抽烟人数高于我们

的估计，这就会使我们的置信区间大于我们所预期的。我们用两个例子来说明这一点。我们使用与前面相同的公式，但是我们这次要求的是 d。如果我们发现抽烟者在人口中的比例是20%，那么这一结果将是：

$$d^2 = \frac{(1.96)(0.2)(0.8)}{323} \qquad d = \sqrt{0.00190} = 0.0436$$

置信区间将在样本估计值的左右两边 ±4.4%，而不是 ±5%。如果我们的调查发现烟民的比例是40%，那么 d 就将是：

$$d^2 = \frac{(1.96)(0.4)(0.6)}{323} \qquad d = \sqrt{0.00285} = 0.0534$$

置信区间不是 ±5%，而是 ±5.3%。

这些例子还说明了另外一个问题。公式有四个组成部分，通过设定或掌握其中的三个部分，我们就可以求得方程中的第四个部分。

公式的假定

我们也应当对公式的其他重要问题予以说明。它是以简单随机抽样(simple random sampling，srs)选取的样本单元为基础的，这就是说所有的单元和单元的组合都有相等的入选机会。实际上，很少有样本是 srs 的，因此公式只不过是对所需样本量的一个参考。其次，公式只使用了一个百分比形式的变量。度量变量(metric variable)如收入，需要使用更为传统的计算方法来计算方差。数据分析或统计课都讲授这种方法。再次，公式给出的只是适合一个变量的样本量。实际情况是，调查中的每一个变量都有自己方差，因而每一个变量需要不同的样本量可能各不相同，它使研究者很难做出抉择。一般讲，研究者最终必须要在样本量、数据收集的方法和研究资源这几者之间有所取舍。

公式的第四个假定是，假定样本来自一个很大的总体，或假定样本只是总体很小的一部分。如果情况确实如此，我们的确可以忽略上面我们求解时使用的 fpc。但是在许多时候，研究者错误地以为，如果样本要能精确地反映总体，它的容量必须占到总体的某一百分比，有时我们常常会听到研究者问这样的问题，是否应该取总体的1%、5%或某个其他百分比的样本。只有在样本代表了5%以上的总体，达到了许多抽样统计学家使用的这一经验标准时，总体的大小才成为一个比较重要的问题。让我们用一个例子来阐述这一问题。我们的烟民研究需要的样本量是323。总体修正因子是由(Cochran，1977)提出的：

$$n' = \frac{n}{1 + \left(\frac{n-1}{N}\right)}$$

我们需要的样本量是 $n = 323$，而 2000 年罗利 18 岁及以上的人口是 $N = 218\ 487$(U.S. Census Bureau, 2000d)。将这些值代入等式，我们得到：

$$n' = \frac{323}{1 + \frac{322}{218\ 487}} = \frac{323}{1.001\ 47} = 322.5$$

给出的解与原来结果的差不到 1，可见二者实际上没有什么差别。如果总体的成年人口为 50 000，那么结果会怎么样呢？我们把这一数字代入等式中的 N，得到的结果是 321。也没有什么差别。我们再来看一个例子。如果总体的人口是 2 500，将这一数字代入等式，得到我们所需的样本量是 286。这比原来要调查的人数少 37 个或 11.5%。注意样本量 323 是总体 2 500 的 12.9%。它已经超过了 5% 的标准。这使我们回到了原来讨论的问题：总体的大小并不对样本的大小发生影响，除非总体很小，且样本超过总体的 5%。正因为如此，如果方差和公式的其他部分假定是相同的，一个在罗利、北卡罗莱纳或整个美国进行的调查，需要的样本量是相同的。

样本量的最大影响因素

现在我们已经对如何确定样本量的问题有了一个初步了解，同时我们知道了在一般情况下，总体大小并非是决定样本大小的因素。那么公式中的哪一或哪些部分对样本量的影响最大呢？表 7.1 列出了方差、概率水平和置信区间的不同组合的样本量。A 部分代表 95% 的置信度（概率水平），而 B 部分则代表 90% 的置信度（概率水平）。第三种结果是，经常使用的概率水平 99%

表 7.1 不同方差（以百分比表示）、概率水平和置信区间需要的样本量[a]

A. 95% 概率水平和不同的方差及置信区间		
$\frac{(1.96)^2(0.5)(0.5)}{(0.05)^2}=384$	$\frac{(1.96)^2(0.5)(0.5)}{(0.04)^2}=600$	$\frac{(1.96)^2(0.5)(0.5)}{(0.02)^2}=2\ 401$
$\frac{(1.96)^2(0.7)(0.3)}{(0.05)^2}=323$	$\frac{(1.96)^2(0.7)(0.3)}{(0.04)^2}=504$	$\frac{(1.96)^2(0.7)(0.3)}{(0.02)^2}=2\ 017$
$\frac{(1.96)^2(0.9)(0.1)}{(0.05)^2}=138$	$\frac{(1.96)^2(0.9)(0.1)}{(0.04)^2}=216$	$\frac{(1.96)^2(0.9)(0.1)}{(0.02)^2}=864$
B. 90% 概率水平和不同的方差及置信区间		
$\frac{(1.64)^2(0.5)(0.5)}{(0.05)^2}=269$	$\frac{(1.64)^2(0.5)(0.5)}{(0.04)^2}=420$	$\frac{(1.64)^2(0.5)(0.5)}{(0.02)^2}=1\ 681$
$\frac{(1.64)^2(0.7)(0.3)}{(0.05)^2}=226$	$\frac{(1.64)^2(0.7)(0.3)}{(0.04)^2}=353$	$\frac{(1.64)^2(0.7)(0.3)}{(0.02)^2}=1\ 412$
$\frac{(1.64)^2(0.9)(0.1)}{(0.05)^2}=98$	$\frac{(1.64)^2(0.9)(0.1)}{(0.04)^2}=151$	$\frac{(1.64)^2(0.9)(0.1)}{(0.02)^2}=605$

a 计算未使用有限总体纠正因子

则留给学生自行解决。表中的每一部分,我们都列出了百分比变量的最大方差(0.50/0.50)、中等大小的方差(0.70/0.30)和较小的方差(0.90/0.10)。同时我们也列出了三种不同的置信区间:±5%、±4%和±2%。

我们先来看一下A部分,它给出的概率水平是95%。一个具有最大方差0.50/0.50的变量和±5%的置信区间,需要的样本量为384。我们再看第一列的下一行,有方差0.70/0.30的变量需要的样本量是323,而方差0.90/0.10的变量需要的是138。注意方差越小,需要的样本量也越小。

在每一行,置信区间则处在变化之中。对于最大的方差和±5%的置信区间,需要的样本量是384。如果我们围绕我们的估计值要求的精度越来越高,那么我们需要的样本量将越来越大。±4%的置信区间需要600,而±2%就跃升至2 401。从±5%到±2%样本量增加了6倍之多。这样的增加同样也发生在方差的变化中。当我们从小方差变为最大方差时,样本量上升了2.77倍,但是我们从一个中等的置信区间变为一个小的置信区间时,样本量上升了6.25倍。且这些变化对不同的方差、置信区间和两个概率水平都大致如此。因此对于一个百分比变量来讲,置信区间的大小或样本估计值的精度,对样本大小的影响最大。当前人口调查便是一个必须对度量变量做高精度估计的样本需求量的实际例子。这一每月进行的调查,必须以小于1%的置信区间估计美国的失业率。为了达到这一百分率,每月要抽取大约56 000左右的住户进行调查。

假设检验和检验力

迄今为止,我们的讨论仅限于那些确定样本量的基本原理。我们感兴趣的问题常常是检验假设或确定样本量的大小和检验力,而非估计总体的百分比。如果我们将我们研究的问题转向这些方面,读者也许会发现几个基本的例子是很有用的。我们先来看几个假设检验方面的例子。假定最近几个全国性的调查表明,约43%登记的选民认为国会的工作做得不错。一个研究者想把自己居住的州的选民的态度与全国选民的态度做一个比较。她对一个200名注册的选民样本进行了调查,并发现她居住的州的51%的选民认为国会干得不错。她想检验一个她的州的选民的态度与全国的没有什么差别的假设。她提出了一个无方向的假设(nondirectional hypothesis)($H_0: p = 0.43 = p_0$ 或 $H_1: p \neq 0.43$),把全美调查的结果作为比较的标准,设 $\alpha = 0.05$,假定正态近似二项分布成立,且在将 p 转化为 z 得分后,再进行有无差别的检验:

$$z = \frac{|\hat{p} - p_0| - 1/(2n)}{\sqrt{\frac{p_0 q_0}{n}}}$$

式中

$\hat{p}$ = 样本估计值

p_0 = 总体值

$q_0 = 1 - p_0$

$\frac{1}{2n}$ = 对二分变量做的连续性修正

$$z = \frac{|0.51 - 0.43| - 0.0025}{\sqrt{\frac{(0.43)(0.57)}{200}}} = 2.21$$

因为 $z > 1.96$,所以我们拒绝 H_0,并认为州的选民的感觉与全国的选民是有差别的。如果样本量 $n = 100$,且研究者调查的结果中持赞赏态度的选民的百分比是相同的,那么这一结果的 z 得分将是 $z = 1.57$。这时,我们就无法拒绝 H_0。而这些不同的结果是由于样本量不同引起的,而检验力问题也就因此而生,我们将在下面一个例子中讨论它。

下一个例子涉及两个随机样本的假设检验问题。假定我们在两个城镇的每一个各对一个 $n = 300$ 的个人样本进行了调查。在城镇一,52%的被调查人说他们是民主党,而在城镇二,则有45%的被调查人说自己是民主党。我们想要确定的问题是,这两个城镇的差别是具有统计显著性的,还是偶然的?与前面的例子类似,我们要检验的假设是 $H_0: p_1 = p_2 = p$ 与 $H_1: p_1 \neq p_2$。我们假定正态近似二项分布成立,假设无方向性,并设 $\alpha = 0.05$,那么公式便是:

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

式中

$\hat{p}_1$ = 从城镇一得到的估计值

$\hat{p}_2$ = 从城镇二得到的估计值

$\hat{p}$ = 从两个城镇得到的结果的加权平均数

$\hat{p} = \frac{n_i\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

$\hat{q} = 1 - \hat{p}$

将我们得到结果代入,得到:

$$z = \frac{|0.52 - 0.45| - \left(\frac{1}{600} + \frac{1}{600}\right)}{\sqrt{(0.485)(0.515)\left(\frac{1}{300} + \frac{1}{300}\right)}} = \frac{0.07 - 0.00333}{\sqrt{0.001665}} = 1.63$$

因为 z 小于 1.96，我们无法拒绝 H_0，我们并因此而得出结论，这两个镇民主党人的比例可能差不多。

有趣的是，如果我们将样本量提高到每镇 $n = 400$，且调查得到的民主党人的比例与原来相同，这时检验的结果将在接受和拒绝 H_0 之间摇摆，究竟接受还是拒绝，要视采用的计算方法而定。将 $n = 400$ 代入公式，计算的结果是：

$$\frac{|0.52 - 0.45| - \left(\frac{1}{800} + \frac{1}{800}\right)}{\sqrt{(0.485)(0.515)\left(\frac{1}{400} + \frac{1}{400}\right)}} = \frac{0.067\,5}{0.035\,34} = 1.91$$

因为这一值小于 1.96，所以我们无法拒绝 H_0。然而，如果我们将分子中的连续性纠正因子去掉，那么结果将是 $z = 0.07/0.035\,34 = 1.98$。这时，我们就可以拒绝 H_0，并据此得出结论，认为两个镇的民主党人的比例是有差别的。这一情况无疑会自然而然地把我们的讨论和例子引向检验力问题。

在我们进行假设检验的时候，我们通常都形成一个零假设或虚无假设（null hypothesis，H_0），并对照一个备择假设（alternative hypothesis，H_1）对它进行检验。大多数统计教科书都对假设检验的步骤和要求的假定做了详细的阐述（参见：Glenberg，1996）。在我们对一个假设进行检验的时候，结果会有四种可能性：

1. 在 H_0 正确时，接受 H_0。
2. 在 H_0 正确时，接受 H_1（甲类错误）。
3. 在 H_1 正确时，接受 H_0（乙类错误）。
4. 在 H_1 正确时，接受 H_1。

因此，我们需要避免发生的错误有两种类型。甲类错误发生的概率 α 是在 H_0 是正确的、不应该被拒绝的时候，拒绝它的概率。在社会科学中，我们一般设 α 或这一错误发生的概率为 0.05。这就是说我们在检验一个假设时，犯错误的概率是 5%，而正确的概率则是 $1 - \alpha$ 或 95%。可见这种类型的错误发生的概率是比较小的。不过，我们可以将概率设得更小，譬如 $\alpha = 0.025$ 或 0.01，但是这将导致第二种类型错误的发生。

乙类错误的概率 β 是在 H_1 正确且应该接受的时候，接受 H_0 的概率。β 和检验力彼此相关。检验力等于 $1 - \beta$，它是在 H_0 应该被拒绝时，拒绝它的概率。这两类错误，α 和 β 彼此反相关。我们使 α 变小，则会导致 β 上升，而减少 β，则会使 α 上升。检验力是一个很重要的统计量，因为低检验力意味着在两个变量之间可能存在一定差异的时候，我们很少有机会能发现有统计显著性的差异。我们为什么给 β 和检验力这两个统计量以这么多的关注呢？主要原因是小容量的样本会导致检验力的低下，因此检验力分析应该作为确定

样本大小的程序的一部分来做。检验力计算的目的在于确定一个适当的样本量,以最大限度地增加在 H_0 应该被拒绝的时候,拒绝它的概率。

尽管影响检验力的方法不止一种(参见 Rosner, 2000; Glenberg, 1996),但在这里我们只想向读者介绍由样本量引起的检验力的变化。我们不妨再一次回到有关登记的选民和他们对国会工作评价的例子。我们准备做一个二项双尾检验,设 $\alpha=0.05$,$n=200$。我们有什么样的方法拒绝 H_0 的检验力水平呢? 拉简(Lachin , 1981)提出了如下的单比例公式:

$$z_\beta = \frac{\sqrt{n}\,|\hat{p} - p_0| - z_{\alpha/2}\sqrt{p_0 q_0}}{\sqrt{\hat{p}\hat{q}}}$$

将我们的值代入公式,得到:

$$z_\beta = \frac{\sqrt{200}\,|0.51-0.43| - 1.96\sqrt{(0.43)(0.57)}}{\sqrt{(0.51)(0.49)}} = \frac{1.13-0.970\,3}{0.499} = 0.322$$

我们到正态分布数值表中寻找 0.322(见表 7.2),发现 $Z_\beta=0.322$ 位于 $Z_\beta=0.39$($\beta=0.35$,检验力 =0.65)和 $Z_\beta=0.25$($\beta=0.40$,检验力 =0.60)之间。这就是说,200 个样本将产生一个检验力在 0.60 ~ 0.65,或为 63% 的检验力。要得到更为确切的数值,我们必须查阅更为详细的统计表,或这一节最后两段列出的若干种信息资源中的某一种。如果我们用样本量 $n=100$ 再一次进行计算,得到的结果是负的,等于 -0.34。如果我们到正态分布表中去寻找这一数值,并把它减去 1,我们发现一个容量为 100 的样本,其产生的检验力为 37%。因此在 $n=100$ 时,我们只有 1/3 的拒绝零假设的可能,或 63% 的犯乙类错误的可能。在这一例子中,用 100 个样本做的假设检验不是一个强力的假设检验。

表 7.2 选定的 α 和 β 值的单位正态偏离

α 和 β	单侧检验或 Z_β 的 Z_α	双侧检验的 $Z_{\alpha/2}$
0.01	2.33	2.57
0.02	2.05	2.33
0.025	1.96	2.24
0.04	1.75	2.05
0.05	1.65	1.96
0.10	1.28	1.65
0.20	0.84	1.28
0.30	0.52	1.04
0.35	0.39	0.93
0.40	0.25	0.84

一条约定俗成的经验法则是将 β 设为 4α。这就是说，如果 $\alpha=0.05$，那么 β 就应当等于 0.20，而检验力就等于 $1-\beta=0.80$。我们使用公式：

$$n=\left[\frac{z_{\alpha/2}\sqrt{p_0q_0}+z_\beta\sqrt{\bar{p}\ \bar{q}}}{\bar{p}-p_0}\right]^2$$

将我们的值代入公式，得到：

$$n=\left[\frac{1.96\sqrt{(0.43)(0.57)}+0.84\sqrt{(0.51)(0.49)}}{0.51-0.43}\right]^2=\left[\frac{0.9703-0.4199}{0.08}\right]^2=302$$

在很多情况下，我们希望能确定双独立样本的样本量。在设计一项干预研究时，一般我们都希望能够这样做。假如我们的研究目的是使烟民人口占 33% 的成年总体的烟民比例降低 30%。假定我们希望实验组和控制组的人数相同（$n_c=n_e$），$\rho_c=0.33$，$\rho_e=0.27$（$=0.70\times0.33$）和 $\bar{p}=0.28$（零假设，它是 ρ_c 和 ρ_e 的均值）。我们将使用 $\alpha=0.05$ 的双侧检验。并令检验力为 0.80。拉简（Lachin，1998）提出的计算总样本量的公式如下所示：

$$n=\left[\frac{z_{\alpha/2}\sqrt{4\bar{p}\ \bar{q}}+z_\beta\sqrt{2\rho_eq_e+2\rho_cq_c}}{|\rho_e-\rho_c|}\right]^2$$

将我们的数值代入公式，得到：

$$n=\left[\frac{1.96\sqrt{4(0.28)(0.72)}+0.84\sqrt{2(0.23)(0.77)+2(0.33)(0.67)}}{|0.33-0.23|}\right]^2$$

$$=\left[\frac{1.76+0.7496}{0.10}\right]^2=630$$

在总样本量为 630 时，实验组和控制组的样本量为各 315 个。假定我们得到的资助只允许我们的样本量为 500。这时，我们的检验力将达到一个什么样的水平呢？我们再次用拉简（Lachin，1998）的公式来计算一下：

$$=\frac{\sqrt{n}\,|p_e-p_c|-z_{\alpha/2}\sqrt{4\bar{p}\ \bar{q}}}{\sqrt{2p_eq_e+2p_cq_c}}$$

$$=\frac{\sqrt{500}\,|0.33-0.23|-1.96\sqrt{4(0.28)(0.72)}}{\sqrt{2(0.23)(0.77)+2(0.33)(0.67)}}$$

$$=\frac{2.236-1.76}{0.892}=0.534$$

在正态分布表中的这一数值表示 70% 的检验力。

我们可以参考查阅的有关样本量和检验力的好资料有许多。亚瑟·格林博格（Arthur Glenberg，1996）编写的《从数据中学习：统计推理导论》（*Learning from Data: An Introduction to Statistical Reasoning*）一书，深入浅出地阐述了单样本和双独立样本的检验力的基本问题。伯纳德·罗斯纳（Bernard Rosner，

2000)撰写的《生物统计原理》(*Fundamentals of Biostatistics*)一书,是这方面的更为高级的一本教科书。该书列举了许多医疗方面的例子,涉及的疾病包括癌症、心脏病、高血压和其他各种疾病。这些例子,在讨论确定样本量和检验力的时候,涉及的内容包括使用单侧或双侧的备择假设、单样本、双样本或纵向样本(longitudinal sample)及二分和距离变量的假设检验和置信区间。

有许多优秀的网站,其中一个是由从乔治城大学退休的生物统计学家约翰·C· 皮兹罗(John C. Pezzullo)管理维护的、链接的站点已经发展到了600个。其网址是:http://members. aol. com/johnp71/javastat. html。这一站点包括了几个链接点,可以帮助我们选择正确的统计方法、免费的软件、书籍和手册、教学指南和许多统计程序的计算方法,要打印这一站点的所有链接需要14页。第一页是一个计算检验力、样本量和实验设计的链接。另外一个非常有用的站点是由拉赛尔 V· 兰斯(Russell V. Lenth)建立的,网址是:http://www. cs. uiowa. edu/ ~ rlenth/Power/。在这一网址可以计算检验力、样本量和置信区间,并为这些方面的问题提供咨询,同时也提供了为数不多的、有用的网站链接。

使用人口普查数据

诚如我们经常提到的那样,人口普查数据在计划一个调查的时候,特别是在估计需要从各种各样的子群体中抽取多少被调查人和这些样本量就分析的目的而言是否足够精确是非常有用的。为了确保有足够精确的样本量,我们必须事先做计划。为此我们必须再一次回顾我们研究的问题,问一问自己,我们研究的目的是什么?我们要提出的问题是什么?我们需要的答案又是什么样的?我们是需要有关整个总体的答案,还是需要同时得到有关整个总体和某些特定的子群体的答案?让我们用罗利成年烟民的例子来继续我们的讨论。

研究目的

我们固然需要考虑一下我们的研究目的,而考虑一下我们认为合适的分析方法同样也很重要。我们在考虑这些问题的时候,不妨查阅一下人口普查的数据,了解一下调查的结果可能会是什么样的。我们仍然从我们研究的问题是什么开始我们的讨论。我们的目的是估计罗利的成年烟民数,因为我们希望建立和检验一个干扰项目,看看它是否能有效地帮助烟民戒烟。很自然,随之而来的问题是我们是否要考虑一些特殊类型的人,即我们是只对估计烟民的总数感兴趣,还是同时也对以性别、年龄、种族或性别和种族组合定义的特殊子群体的烟民数感兴趣?我们之所以问这些问题,是因为研究告诉我们,抽样行为因不同的子群体而异,因此如果我们根据不同子群体的不同情况推

出不同的干扰项目,而非推出一个适用于每一个人的普遍性项目,戒烟的效果会更佳。由此可见,我们感兴趣的并非只是整个总体的估计值,同时也包括各个特殊子群体的估计值。

分析组

随后,我们应该问一问我们自己我们感兴趣的子群体是什么样的。假定我们对以男女分组、三个年龄段分组(18～39、40～64 和 65 岁及以上分组)、种族分组,也许是性别和种族的交互分组的估计值感兴趣。在这一点确定之后,接着我们就要弄清楚罗利市这些群体的人数,确定我们可分别为我们的调查提供多少个调查对象,然后再把来自不同群体的信息加总在一起。最后一步使我们得以估计所有子群体的置信区间,并确定其精确度是否能满足我们的研究目的。

假定我们的资源允许我们做一个 600 个调查对象的、时间在 30 分钟之内的电话调查。为了估算一下这些被调查对象应该怎样在我们感兴趣的子群体之间进行分配,我们需要找到描述我们感兴趣的特征的罗利成人分布的数据,并做出一些有关电话调查是否精确地代表这些群体的假定。

查阅普查数据

而这时,查阅一下人口普查数据可谓正当其时。美国人口普查局进行的各种普查,包括人口和住房普查、农业普查和政府普查等。经济普查则涵盖零售、批发、服务业、交通运输、制造业、矿业和建筑业。此外还有有关外贸和其他领域的各种普查。普查通常 5 或 10 年进行一次,间隔时间的长短视调查的内容而定。调查数据代表了两种类型的地区:行政区域——美国、波多黎各、州、县、城市、乡村、镇区、选举区、印地安人保留区等;统计区域——美国地区(regions of the United States)、大都市和小都市区(metropolitan and micropolitan areas)、都市区、普查地域(census tracts)、街区或街区群、邮政编码区(ZIP codes)和其他类型的地区。在 http://factfinder. census. gov 网站,可在线获得有关数据。数据形式包括打印的报告、光盘和微缩胶卷等。除了这些普查之外,美国人口普查局每年还要进行 250 个左右的抽样调查。总之,美国人口普查局在挖掘、收集和发布各种珍贵且很有用的信息。

就我们的罗利烟民调查而言,我们感兴趣的数据来自人口和住房普查。这一调查每 10 年进行一次,它从每一个住户那里收集有关的信息。问卷分长短两种,短问卷的内容包括性别、种族、年龄、婚姻状况和居住单位内所有成员在户中的关系。长问卷在样本中散发,增加了一些额外的问题,包括社会和经济特征、受教育程度、出生地、血缘家系、在家使用的语言、残疾、生育、就业状况和一些其他问题等。

表 7.3 别年龄、性别、种族和西班牙血统的罗利 2000 年人口构成表

地区和[入选州的]县部分[人数在 10 000 或 10 000 以上]	罗利市						
	总人数	白人	黑人	美籍印第安人、爱斯基摩人或阿留申人	亚洲或太平洋岛民	西班牙血统（任何种族）	白人、非西班牙血统
总人数	276 093	178 649	78 844	2 091	10 872	19 308	166 386
5 年以下	17 461	9 633	6 469	142	788	2 043	8 177
1 年以下	3 768	2 158	1 309	36	146	523	1 798
1 年	3 653	2 021	1 329	24	195	445	1 698
2 年	3 433	1 867	1 306	26	140	407	1 590
3 年	3 328	1 825	1 236	26	150	354	1 570
4 年	3 279	1 762	1 289	30	157	314	1 521
5 ~ 9 年	16 444	8 711	6 702	127	683	1 304	7 726
5 年	3 314	1 785	1 310	21	130	315	1 544
6 年	3 348	1 768	1 383	25	140	271	1 558
7 年	3 255	1 698	1 329	29	142	264	1 514
8 年	3 299	1 746	1 354	16	144	236	1 565
9 年	3 228	1 714	1 326	36	127	218	1 545
10 到 14 年	15 254	8 405	5 963	128	616	903	7 725
10 年	3 290	1 780	1 325	24	129	189	1 632
11 年	3 191	1 747	1 267	24	130	198	1 594
12 年	3 030	1 669	1 184	28	123	171	1 541
13 年	2 865	1 605	1 098	25	104	190	1 474
14 年	2 878	1 604	1 089	27	130	155	1 484
15 到 19 年	19 864	12 355	5 934	187	801	1 924	11 169
15 年	2 805	1 613	1 007	19	130	179	1 478
16 年	2 736	1 565	989	29	97	254	1 391

续表

地区和[入选州的]县部分[人数在 10 000 或 10 000 以上]	总人数	白人	黑人	美籍印第安人、爱斯基摩人或阿留申人	亚洲或太平洋岛民	西班牙血统(任何种族)	白人、非西班牙血统
罗利市							
17 年	2 906	1 649	988	32	137	353	1 413
18 年	4 772	3 038	1 334	51	183	513	2 742
19 年	6 645	4 490	1 616	56	254	625	4 145
20 到 24 岁	32 458	21 072	8 213	308	1 437	3 763	19 036
20 岁	6 725	4 545	1 590	64	239	706	4 173
21 岁	6 616	4 351	1 672	72	277	741	3 926
22 岁	6 538	4 267	1 667	56	281	741	3 845
23 岁	6 340	3 998	1 668	59	311	786	3 590
24 岁	6 239	3 911	1 616	57	329	789	3 502
25 到 29 岁	30 850	19 591	8 138	246	1 568	3 281	17 763
30 到 34 岁	26 255	16 436	7 588	210	1 226	2 251	24 004
35 到 39 岁	23 320	14 866	6 908	185	939	1 495	21 825
40 到 44 岁	20 506	13 334	5 976	170	821	885	19 621
45 到 49 岁	17 950	12 240	4 853	108	658	571	17 379
50 到 54 岁	15 034	10 722	3 694	97	461	359	14 675
55 到 59 岁	10 308	7 657	2 282	48	307	204	10 104
60 到 64 岁	7 394	5 561	1 605	40	195	121	7 273
65 到 69 岁	6 246	4 707	1 350	26	171	75	6 171
70 到 74 岁	5 779	4 522	1 150	24	90	52	5 727
75 到 79 岁	4 957	3 939	947	15	63	39	4 918
80 到 84 岁	3 186	2 600	556	20	29	21	3 165
85 岁以上	2 827	2 298	516	10	19	17	2 810
18 岁及以上	218 487	147 073	56 726	1 614	8 421	14 272	52 570
62 岁及以上	27 249	21 275	5 442	115	482	259	26 990
65 岁及以上	22 995	18 066	4 519	95	372	204	22 791

年龄中位数	30.9	32.7	28.8	28	28.6	24.6	31.8
女性	139 445	89 443	42 219	1 019	5 183	7 339	84 392
5 年以下	8 548	4 668	3 224	69	391	977	3 973
1 年以下	1 845	1 044	658	19	74	256	863
1 年	1 798	990	666	11	98	211	828
2 年	1 678	912	642	12	74	187	786
3 年	1 617	867	612	11	71	177	751
4 年	1 610	855	646	16	74	146	745
5 到 9 年	8 064	4 271	3 278	68	355	627	3 792
5 年	1 604	848	643	14	78	142	731
6 年	1 653	854	689	11	75	149	741
7 年	1 617	849	651	12	77	130	758
8 年	1 589	848	650	10	60	111	770
9 年	1 601	872	645	21	65	95	792
10 到 14 年	7 563	4 095	3 020	75	310	448	3 759
10 年	1 627	861	682	14	61	95	778
11 年	1 550	828	640	12	64	85	760
12 年	1 459	803	567	16	59	87	745
13 年	1 467	822	557	16	64	101	754
14 年	1 460	781	574	17	72	80	722
15 到 19 年	9 798	6 005	3 137	84	368	726	5 557
15 年	1 382	782	519	10	63	84	713
16 年	1 342	751	506	11	50	104	680
17 年	1 408	786	506	14	67	119	706
18 年	2 410	1 535	721	23	78	183	1 430
19 年	3 256	2 151	885	26	110	236	2 028
20 到 24 岁	15 474	9 923	4 410	149	625	1 241	9 182
20 岁	3 235	2 173	853	28	103	249	2 028

续表

地区和[入选州的]县部分[人数在10 000或10 000以上]	罗利市						
	总人数	白人	黑人	美籍印第安人、爱斯基摩人或阿留申人	亚洲或太平洋岛民	西班牙血统（任何种族）	白人、非西班牙血统
21岁	3 106	2 006	884	35	130	232	1 855
25到29岁	14 617	9 073	4 354	100	734	1 063	8 417
30到34岁	12 672	7 740	4 102	83	573	757	11 915
35到39岁	11 434	7 153	3 655	94	431	520	10 914
40到44岁	10 399	6 625	3 228	86	409	335	10 064
45到49岁	9 342	6 251	2 688	60	345	217	9 125
50到54岁	7 901	5 613	2 007	43	227	162	7 739
55到59岁	5 412	3 970	1 265	26	148	92	5 320
60到64岁	4 015	3 001	899	25	95	65	3 950
65到69岁	3 491	2 591	804	15	88	40	3 451
70到74岁	3 395	2 613	732	8	44	29	3 366
75到79岁	3 012	2 354	623	11	32	17	2 995
80到84岁	2 138	1 741	384	17	13	11	2 127
85岁及以上	2 170	1 756	409	6	13	12	2 158
18岁及以上	111 138	74 090	31 166	772	3 947	4 980	106 158
62岁及以上	16 490	12 757	6 470	72	245	137	16 353
65岁及以上	14 206	11 055	2 952	57	190	109	14 097
年龄中位数	32.1	34.2	29.6	27.9	28.7	23.6	32.8
男性	136 648	89 206	36 625	1 072	5 689	11 969	124 679
年龄中位数	30.0	31.4	27.7	28.1	28.3	25.1	30.8
每百名女性中的男性	98.0	99.7	86.8	105.2	109.8	163.1	94.4

我们可从网址 http://factfinder. census. gov 2000 年普查的摘要报告中(Census 2000 Summary File 1,SF 1)得到我们需要的信息。表 7. 3 提供的数据是罗利市的特定年龄和年龄组的、分种族和西班牙血统及单性别组——女性的总人数。要了解分年龄和分种族的男性人数,我们只要用对应的总人数减去女性的人数。只要稍微再做一些额外的工作,减法和除法,我们就可以从这张表中得到所有我们需要的信息。

做选择

表 7. 4 为我们提供了 600 个样本的、分子群体的最终应该完成的调查人数的估计值。在构建这一表格之前,我们先来查阅一下住房普查的数据(U. S. Census Bureau, 2000a),并由此得悉罗利已经有人居住的住房中,只有 1. 6% 户中没有电话。因此,我们假定,电话在总体的不同子群体中的分布是相当均匀的,因此没有一个我们感兴趣的子群体在样本中可能显著不足。我

表 7. 4 罗利市 600 个样本的电话调查选定的特征子群体的调查人数估计

	总体		
	N	%	估计调查人数
种族			
白人	147 073	67. 3	404
黑人	56 726	26. 0	156
其他人	14 688	6. 7	40
	218 487	100. 0	600
性别			
男	107 349	49. 1	295
女	111 138	50. 9	305
年龄			
18 ~ 39	124 300	56. 9	341
40 ~ 64	71 192	32. 6	196
65 岁及以上	22 995	10. 5	63
18 ~ 34	100 980	46. 2	277
35 ~ 64	94 512	43. 3	260
65 岁及以上	22 995	10. 5	63
18 ~ 34	100 980	46. 2	277
35 ~ 59	87 118	39. 9	239
65 岁及以上	30 389	13. 9	83
种族及性别			
白人男性	72 983	33. 4	200
白人女性	74 090	33. 9	203
黑人男性	25 560	11. 7	70
黑人女性	31 166	14. 3	86

们假定电话调查将以相同的比例通达这些子群体中的每一个,即在每一个子群体中,电话分布的比例是相同的。这就是说,如果我们的样本的确是随机的,且各个子群体的调查合作率也是相同的,那么我们最终得到的调查结果的百分比,将会与总体十分相似。我们并不期待我们的数值与表 7.4 所列的完全匹配,但是我们假定它们将与这些估计值大致相当。

表 7.4 显示了人口中常见的分布情况,性别分布在男女两性之间非常平衡。如果调查的合作率在两性间也大致相等的话,那么男女的样本量也应该相等。然而,在大多数调查中,这一假定是否正确似乎是很值得怀疑的,因为女性比男性一般更为合作。在最后的调查结果中,男性的百分比通常总是比总体的低几个百分点。如果像本例那样,男性占据了合格总人口中的 49%,但在最后的访谈结果中,它可能在 42% ~46%。而现在,我们不想急于用估计的合作率对表中的数字进行修正,就用表中的数字作为我们分组结果的最佳估计数。我们需要回答的问题是:"我们将要在每一个子群体调查的人数,是否满足了我们的研究目的?"实际上,这个问题就是我们是否能够接受各种各样的样本量产生的精度。

我们先来做一些计算,然后看看结果如何。我们预计将要调查 295 个男性。假如我们设定的概率水平是 95%,方差为 0.30/0.70,要求的统计量是 d,我们求得的结果是 ±5.2%。我们怎样做,一个 10.4% 的置信区间(值域)才是合理的呢?

我们先来看一下种族。我们的调查将要调查的白人是 404 个,黑人是 156 个,而其他种族的只有 40 个。这样就会存在两个比较严重的不平衡问题。一是 40 个被调查对象不论对何种群体,要做一些有意义的分析其数量都太少。我们可能有的选择是:①将另一个群体与白人或黑人合并;②在调查开始的时候,询问被调查对象的种族,将其他种族的调查对象从调查中筛除,并将那些既非白人也非黑人的人视为不合格的调查对象;③调查所有的被调查人,将其他种族的被调查对象排除在分析之外。选择②和③似乎更好一些,但是这并非轻易就可以做的决定。将其他种族和白人或黑人合并的策略是不可取的。因为情况很可能是,且有文献为证,华人、日本人、马来人和其他血统的人,对抽烟所持的态度不仅彼此很不相同,而且与白人和黑人也很不相同。因此,对于任何这样一个群体来讲,要做一些有意义的分析,访谈的人数都太少了。选择②多少有一定的风险。因为为了筛除 40 个不合格的调查对象,而用一个有关调查对象的种族非常个人化的问题开始我们的调查,去冒得罪我们可能的调查对象和拒访率上升的风险是不值得的。而选择③将我们的访谈数据弃置不用,这似乎是一种不智之举。更何况我们的预算本来就很紧。最佳的策略也许是在试调查的时候,对选择②加以测试。如果筛选的确影响了拒访率和/或影响了费用,这时选择③可能不失为一种可以接受的折中策略。

第二种不平衡与白人和黑人的调查对象人数差别很大有关。我们前面有关样本容量的那些例子已经指出,每当群体组间方差相等,但样本量不同的时候,置信区间的大小也将有所不同。理论上讲,在要进行组间比较的时候,我

们希望各组的置信区间相等。使用表7.4列出的样本量的估计数和用于其他种族的假定,白人和黑人的置信区间分别为±4.5%和±7.2%。区间范围的大小,白人是9%,黑人是14.4%。对于大多数社会科学的变量而言,这样的范围都可以看作是精确的,而对于其他学科,需要的精度可能要更高一些。随着百分比越来越接近0.50/0.50,这些范围也变得越来越大。而一种可供我们选择的方式是多抽取一些黑人。这就意味着,我们要对住户的种族进行筛选,而这种做法诚如我们前面所言,其结果可能是负数。筛选也会增加费用,降低我们可以支持的调查人数。做出正确的选择并非一件容易的事。在我们遇到类似这样的情况时,研究者必须小心谨慎地看一看那些关键的变量和在做出最后的设计决策时需要达到的精确度。

请大家回想一下每一个变量或每一个变量的组合都有它自己的方差这一点。如果我们对白人男女和黑人男女的态度和行为感兴趣,那么我们就必须做同样的计算。表7.4显示,我们可以预期在每一个白人的性别分组中都有200个被调查对象,而黑人的只有70~86个。我们可以从我们以前做的计算中,得到白人组的置信区间的大致概念,这样就让我们来看一下黑人男性组,即最小的那一组的置信区间可能会是什么样的。采用0.30/0.70的方差和95%的概率水平,样本量70给出的置信区间是±10.7%,或范围为21.4%。如果我们的调查发现30%的黑人男性是烟民,那么95%的概率水平提供的总体中黑人男性的烟民比例的估计值在19.3%~40.7%。如果这一变量是一个关键的变量,那么我们就要考虑增加样本量的问题了。还有一点我们希望能给读者交代一下。如果我们对年轻的黑人男性烟民的行为感兴趣,那么我们得到的调查对象人数甚至会更少,样本容量也会因此而很不精确。

表7.4列出的年龄的数字显示了另外一个重要的问题。最初,我们感兴趣的三个年龄组是:18~39(青年)、40~64(中年)和65及以上(老年)。这些年龄组的人数在总体中并不相等。大多数调查对象将来自青年组,而老年组的人数最少,甚至比黑人男青年还少。另一种可供我们选择的、能较好的实现各组的样本量大致相等的方法是改变我们年龄组的定义。当然,只有在有实际意义时,我们才这么做。例如,我们可以将青年定义为18~34,而将中年定义为35~64。这样做使每一组的置信区间比较接近,但却没有改变老年组的情形。另一种可能是重新定义中年和老年,即将中年定义为35~59,而将老年定义为60岁及以上。这一改变,使老年的样本量有所改观,从而使它优于黑人男性青年。当然,采用其他的定义也同样是可能的。

我们在结论中将再次强调,那些能使我们的研究具有最高效率的数据是可以得到的。即使我们有一个极佳的研究思路,我们也不应该盲目地启动我们的研究和在毫无计划的情况下开始收集数据。这种做法无疑会无法使我们的研究得到最好的可能结果。例如,在决定研究白人和非裔美国人的态度之后,未做周密的计划就进行了调查,但后来却发现非裔美国人只占总体的12%,我们得到的样本量根本无法对他们做精确的分析。这样的做法是完全不可原谅的。老年——年龄在65岁及以上的人,通常在总人口中只占很小的

比例。在开始设计老年研究之前,不仅应该对这一比例有所了解,而且也应该对老年人的回答率一般比其他年龄组的人要低这一点有所了解。

普查数据、来自类似总体的现有研究、教科书、数据文件文档和其他各种来源都可以为我们提供许多有用的信息,帮助我们设计和计划一项高效的研究。本书第10章和参考书目也提供了许多信息源。我们在设计或实施一项研究时,可以去那里进行查询。

注 释

1 10年一度的人口普查数据还有一个缺点,就是在数年之后它就会过时,在那些高度发展的地区,情况尤其如此。克服这一缺点和人口调查员的困难的方法是建议进行美国社区调查。这样的调查每年进行一次,可以详细地收集人口、社会经济和住房方面的数据。它将提供年度的和多年的这些方面情况的估计数,并将完全取代10年一度的人口普查的长卷。2000年人口普查和同一时间进行的大规模的补充调查的数据质量的比较,可以在美国人口普查局(U. S. Census Bureau,2004)找到。

2 抽样误差的含义并不等同一般所谓的犯错误。它是一个估计的值域,由我们的样本计算得来。它表明总体值可能在什么地方。估计的值域是计算得到的,因为我们调查的只是总体的一个样本,而非总体中的每一个人。

3 为什么每一个街区要调查5个人,这一个问题将在第8章进行讨论。

4 在抽样的第一阶段,用与大小成比例的概率选取地方这一方法,给比较大的地方以比较高的入选样本的机会。例如,我们打算在100个县中选取24个。在北卡罗莱那,2000年,18岁及以上的人口数量是6 086 266。我们用6 086 266除以24,得到抽样间距为253 553。奥斯陆县18岁及以上的人口数量是111 017,而阿希是19 557。在第一阶段,奥斯陆入选的概率(111 017/253 553 =0.438)高于阿希(19 557/253 553 =0.077)。为了使整个的选择概率相等,在最后一阶的抽样时,抽取单位数必须相等。这一点将在第8章阐述。

5 加权问题将在本章后面的部分及第8章和第10章讨论。

6 这一点将在第8章介绍。

7 这一信息是我们通过电话从美国通信委员会(FCC)的行业分析处得到的,它不包括那些既有住宅电话又有营业电话的住户。

8 样本单元可以回置和无回置地抽取。回置抽取的方法是:如果我们打算从1~10这一组数字中,随机地选取3个数字的样本,那么在每选出一个数字之后,我们再把它放回原来的数字组,而在随后的抽取中,它可能再一次被抽到。

9 当我们用这一例子一步一步地去做的时候,读者势必会考虑如何将这些步骤用于不同的例子,如用于估计自己家乡每周至少在家用5次早餐,或每周至少运动3次,每次运动30分钟的成人比例。这些备择的例子无疑会加深我们对这些方法步骤的理解。

第8章 选取样本

Selecting a Sample

电话、邮寄和互联网调查是学生最有可能做的三种类型的调查。在这一章介绍的许多决策方法是我们在为地区性电话调查时选取样本所必须做出的，这些方法还可进而推广到全国性的用随机拨号(RDD)法选取样本。然后，我们将用两个例子来向读者介绍如何用清单来抽样。这样的清单应该是：①清单中个体名单排列有序，可用来选取分层的样本；②清单的结构可用于选取由一个个自然的群体组成的样本，如由一班学生组成的样本。

表8.1 抽样步骤

• 定义总体
• 寻找和/或构建抽样框
• 确定样本量
• 定义总体
• 抽样
• 控管样本和再抽样
• 选取被调查人

表8.1列出了抽样的主要步骤。本章列举的各种例子对这些步骤进行了具体的阐述。在我们开始介绍第一个例子之前，先向大家介绍一下这些步骤，以使大家能对我们阐述这些例子的基本思路有所了解。

例一：社区性的清单辅助的电话调查样本

定义总体

在开始讨论这一问题之前，我们先来回顾一下一项调查组织实施的最初几个阶段。假定我们希望研究罗利地区年龄在18岁及以上的成年人对流产问题所持的态度。我们已经得悉，用电话调查来收集有关资料是可行的。因为这种方法不仅回答效应(response effects)较低，且和我们的工作日程没有冲突，且费用也在预算范围之内。回答效应是一种由回答问题所产生的各种误差。它是由记忆不完整、被调查人错误地理解了问题，或数据收集方法不当所

引起的(Sudman & Bradburn, 1982)。我们也已经了解了没有电话的住户的比例,且对这一问题可能产生的影响也已经有所考虑,并愿意接受由此而产生的偏倚。在威克县和罗利市,这一数字估计分别为1.2%和1.6%(U.S. Census Bureau, 2000a),都低于全国的平均数。

选择抽样框

我们下一个任务是找到一个可以接受的抽样框,并确定它的覆盖范围和新旧程度。罗利市的电话簿上的部分居民估计只登录了住户的电话,并不包括营业性的电话。这个电话簿包括罗利(2000年的人口=276 093)、卡里(2000年的人口=94 536)、6个市镇(2000年的人口=57 428)和农村及行政区划未定地区(2000年的人口=87 264)。县的人口的大部分都包含在了电话簿中。因为自由主义和保守主义对流产问题的态度可能因居住地而异,加之研究本身也希望对这一假设和其他假设进行检验,所以电话簿的覆盖范围基本适合我们的研究目的。

然而,作为一个抽样框,电话簿存在一个很大的不足,那就是它不包括那些未曾登录的电话号码。据国际抽样调查公司估计(Piekarski,1997),在罗利—达拉谟—教堂山都市统计区(Raleigh-Durham-Chapel Hill metropolitan statistical area)未登录的电话号码的比例约为21.8%。这样的潜在偏倚已经大到不能被忽视的程度,因此,我们的调查需要使用RDD样本。

现在为大多数调查组织和政府机构所青睐的RDD法是以清单为辅助手段的RDD法。这一方法首先由卡沙迪和列普克斯基(Casady & Lepkowski, 1993)提出和检验的。它是作为一种以清单辅助的分层设计提出的。博瑞克、沃克斯博格、卡尔普和斯达勒(Brick, Waksberg, Kulp, & Starer,1995)及塔克、列普克斯基和皮尔卡斯基(Tucker, Lepkowski, & Piekarski , 2002)已经证明这种设计可以简化用于简单随机样本(simple random sample,srs)的设计,且偏倚只是略有上升而已。srs设计是使用最为普遍的方法,特别是在州、区域和全国性样本的设计中,它的使用尤为普遍。与多阶抽样设计相比,srs具有保存记录和分析方法更简单和费用较低这两个优点。

清单辅助设计的关键是得到一张包含所有前置号码的清单及若干至少含一个已经登录的住宅电话的百位号码库(hundred bank)字段。一个百位字段是未包括地区码的七位电话号码的前五位。换句话说,它是前置号码加上四位后置号码的前两位。七位电话号码的最后两位就是百位号码库,因为这些数字由100个可能的号码组成:00—99。只要得到了一张五位数的号码清单,我们就可以先从这些号码中选取一个系统或简单随机样本,然后再在样本中的每一个号码后面随机地加上一个两位数,生成一个七位数的电话号码的随机样本。我们将在下面的例二中,讨论选取多少个号码和具体的选择方法。

前置号码和至少含一个住宅号码的百位号码库可以从若干商业机构那里得到。海尼斯和科公司备有 75 个主要市场的电话簿。其中包括 9 个州和哥伦比亚特区。电话簿刊登了分街道地址或以号数顺序排列的登录的电话号码。后一种电话簿可叫做电话用户信息（telokey），它正是清单辅助的样本所需要的那种电话簿。另一个信息源是美国信息（InfoUSA）。他们出售电话簿和其他的公共信息资料，也对这些资料进行编辑整理，然后为盈利或其他目的出售。我们可以从他们那里买到分地区码、区域或其他我们感兴趣的地理分布的七位电话号码的全部清单。

购买电话号码的全部清单，并自己动手进行抽样，特别是在一个相当大的总体范围，这种做法有一个不足，即费用太高。如果清单只使用一次，而不是用于若干次调查，这时购买一个清单辅助样本的费用，一般都比购买全部清单低廉。大多数调查团体都与国际抽样调查公司和市场营销系统（Marketing Systems Group/Genesys）这两家公司共事。

例二：基于电话簿的区域性电话样本

在一个由一本电话簿就可代表的比较小的地理区域内做调查的时候，比较简单易行的方法是选取一个基于电话簿的 RDD 样本。在这个例子中，我们将从电话簿中随机地选取号码，去掉号码中的最后两位，然后以随机选取的两位数代替它们。这一方法隐含两个假定。一个是电话簿已经包括了所有的前置号码，且自出版以来没有新增任何前置号码。新增的前置号有可能导致偏倚，因为那些分配到新的前置号的住户，将不可能进入我们的研究。我们假定这一电话簿只是在几个月之前才出版，因此这样的可能不存在。然而，在对这一问题遇有疑问时，我们应该打电话到电话公司询问。第二个假定是，未登录的电话号码是在登录的电话号码中按比例分配的。如果这一假定不真实，且未登录的号码集中在少数几个前置号，那么未登录的电话号码入选样本的概率可能比较小。另一个应该关注的问题是，在许多地理区域，前置号限于某些特定的地理位置。如果一个区域未登录的电话号码是不成比例的，那么它在样本中的代表不是过多就是过少。而这样的问题一旦发生，就会导致最终结果的偏倚。我们不知道任何曾经对这一假定进行过检验的研究（确切地讲，我们无法对这一假定进行检验——译者注），研究者要么假定分布是成比例的，要么对这一潜在问题忽略不计。

在这一研究中，我们将设计选取一个系统的随机样本，选取将分阶段进行。为了实现这一目标，我们需要估计样本选取的总数，计算抽样间隔和随机起始点。这一方法不仅要比 srs 法简单许多，而且更有效率。例如，我们假定

罗利电话簿中住宅电话部分共有 440 页,每页有 4 列,每列共有 104 行。为了从这一本电话簿中选取一个简单随机样本,我们必须给刊登在 440 页中的每一个住宅电话号码一个编号,这样总的编号数约为 122 500 个,不仅如此,我们还需从 1 ~ 122 500 这么多编号中随机地选取唯一的编号。困难之处在于最初对全部电话号码进行编号。你们不妨设想一下,写出 1 ~ 122 500 之间的全部编号,可能需要多少时间[1]?

表达式 8.1 是为计划进行的电话调查确定所需选取的电话号码数的公式。在随后的若干个段落中,我们将要对它进行一些讨论:①如何求解;②表达式隐含的假定;③如何得到我们将要用于求解的数目和百分比。因为表达式或表达式中的某些部分将要在这一章中被用到若干次,我们将尽可能详细地对它进行一些讨论。表达式的文字含义是:为了确定样本选择的数目,必须将计划最终需要完成的调查数目,除以在抽样的每一阶段合格的电话号码的百分比和最终完成调查的住宅号码的百分比的乘积。

$$\text{抽取的样本数} = \frac{\text{需要完成的调查数}}{\left(\begin{array}{l}\text{电话簿中}\\\text{住宅电话}\\\text{的百分比}\end{array}\right) \times \left(\begin{array}{l}\text{RDD 拨号中}\\\text{住宅电话的}\\\text{百分比}\end{array}\right) \times \left(\begin{array}{l}\text{住宅电话中}\\\text{最终完成调}\\\text{查的号码的}\\\text{百分比}\end{array}\right)} \qquad (8.1)$$

分母中的三个部分与抽样和数据收集过程中的三个阶段相对应,那时,不合格的被调查人或未果的调查时有发生。抽样率必须对不合格的调查对象和未果的调查做出弥补。

求抽样数

为了确定样本所需选取的数目,我们需要考虑整个样本选取的过程,并对调查员的工作质量有所估计。首先,我们需要确定最终需要完成的调查数。假定我们最终需要调查 500 个。然后,我们需要考虑整个样本选取的过程。我们将要从一本假设的电话簿中选取住宅电话号码并将去除抽出的电话号码中的最后两位,代之以两位随机数。我们必须问的一个问题是,我们使用的样本框质量究竟如何?是否每一张列表都有相同选择的可能性。为了回答这些问题,我们必须检查样本框。

表 8.2 是电话簿一部分的一个摹本。我们来看一下第 2 列,并非打印的每一行的结尾都是一个电话号码。例如大家注意一下,B. Edward Smith's 一户的资料就占据了三行:第一行列出了姓名,第二行列出了营业地址和电话号码,而第三行才是住址和电话号码。我们可以看到我们的抽样框并不那么“纯粹“——其中有一些是营业性的电话,此外,虽然有些人只列了一行,但有些人则列了几行。一个人在电话簿上占据的行数越多,他被选择的概率就越

表8.2 一页电话簿上的部分内容的摹本

SMIT–SMITH *AREA CODE 101* *RESIDENTIAL SECTION* *378*

SMIT Robert 231-0401
SMITE A J 276-1112

SMITH—See Also Smythe

A B 8403 Lawrence Rd 482-1364
A Curt 113 Adam Ln Ashe 221-0123
A Don 804 Monk Cir Mdvl 253-4444
A Ed 7821 Adam Ln Ashe 451-2233
A Floyd 831 Woodley Rd 785-4131
A Gilbert 7740 Lawrence Rd 482-9264
A Helen 144 Wolfline Ashe 451-7203
A Irene 1113 Winslow Ln 281-4455
A Jess 4455 Hillcrest Rd 785-9999
Anthony K 123 Lyons Rd 223-7845
A Paul & Wanda
322 Adam Ln Ashe 451-6547
A Robert 742 Smallwood Ln 271-2801
.
.
.

SMITH Brock 1775 Charles St. 271-1661
B Charles 784 Falls Rd 221-7743
B Douglas High St Ashe 221-0909
B Edward
Ofc 154 Flint St 826-1122
Res 1225 Charles St 271-6665
B Frank 621 Beaver Ln 463-7723
B Greg & Michelle P
6698 Prince Charles Rd 271-3434
B Harry 686 Round St Mdvl 253-7246
B Irv 9465-C Clarke Ln 785-5232
B John 81 Mullen St 281-8552
B Kelly 4653 Bryant Cir 482-0013
B Lawrence
243-B6 Apple Blossom Ct 482-0065
B Ryan 781 Main St Ashe 221-6251
B Steve 200 Newport Ct 787-7831
B Thomas 154 Styles Rd 223-4654
.
.
.

SMITH C Allen 272 Lane St 271-0408
C Bruce 248 Brooks Ave 271-7003
C Dallas CPA 10 Main St 785-2838
Res RR 1 253-1180
CF 7812-E Apple Blossom Ct. .. 482-6001
C George & Margaret Ann
5720 Benton Rd 785-9902
C Hal 421 Davis Rd 271-7801
C Ike 1013 Falls Lake Ashe 221-2311
C John 23 Wilber Hall 281-0028
C Kurt Ofc 10 W Main St 271-6050
486 Brentwood Mdvl 253-6112
C Luke 333 Weaver Ln 785-2331
C Molly 2517 Country Club ... 463-4545
C Nick 6442 Clarke Ln 785-6668
C Quinton 510 Adam Ln Ashe .. 221-9155
C Rick 954 Bryant Cir 482-6835
C Scot 84 Prince Charles Rd ... 271-0003
.
.
.

SMITH D A 452 W Lane St..... 463-3629
D Billy & Sally 172 Lane St... 271-9103
D Carl 3334 Vernon 421-4746
Daryl D 4127 Shela Ln 483-1988
D Elvis 333 Drexel Ave...... 271-0369
D Francis 6235 N Green St ... 785-1111
D Glen 1534 Lakes Rd 482-4371
D Hank 638 Lewis St 482-5300
D Islam 3947C Pepper St..... 281-4400
D Justin 1221 Shellie Rd..... 482-2200
D Kirk 2604 W Green St..... 276-1662
D Linda 767 Sherron St...... 483-1933
D Mark 1517 Burn St 271-0839
D Neil 516 Sanders Ct 271-5915
D Opra 946 Harvey St. Ashe.. 221-0009
D Phillip 2854 N Clark...... 783-5815
D Raymond 1735 Harris St... 751-0435
D Yales 12 E Lane St........ 251-6666
.
.
.

注:姓名、地址和电话号码是虚构的。

高，或进入样本的可能性就越大。为了均等选择的概率和估计合格的行数，我们必须确定一条决策的原则。我们的决策原则是，只有在抽出的行含有住宅号码时，我们才决定选取这一电话号码。例如，第三列中列有 C. George 和 Margaret Ann Smith 两人姓名的那一行，将不进入我们的样本，并把它看作一个空行，作为不合格处理。然而，如果我们抽到了下面一行，那一行列着一个住宅电话号码，我们就应该把它取出，并放入我们的样本。

我们也必须估计列着营业电话的行和空行的比例。我们通过从电话簿中选取一个样本来估计这一数字。在这一调查中，我们先随机地在 440 页中选取了 25 列，然后再计数每一列的空行和营业性电话数。计数结果的总数为 231，也就是说总行数的 8.9% 是空行和营业性电话[2]。我们很快便会用到这一信息。

我们在用替代的数字创立了新的电话号码之后，必须周详地考虑这一号码的可能性。基本的可能性有三个：是一个合格的住户号码、是一个不合格的住户（如一个营业场所）号码，或号码的性质不得而知。我们先来讨论后两种状况。

在我们创立一个新的电话号码的时候，它可能是一个营业性的号码、一个未曾接通的电话号码或一个不再运行的号码，而所有这三种情况都可归结为不合格这一类。在计算抽样数的时候，我们必须估计在选取的号码中可能是不合格号码的比例。而估计这一数字的最好方法是查阅以前的调查。然而，在我们的调查是在这一地区第一个使用这一方法时，必须做一个试调查。我们可以用试调查得到的结果来指导我们以后进行的正式抽样。

另外一个可能也会对选取的样本数发生影响的因素是估计的未知状态的号码数。这些号码可以接通，但是只有电话铃声响，从来没有人接听。许多研究者都要求调查员必须在挂断之前，让电话铃声响至少七次。然而，有些号码在一天的不同时间和每星期的不同时间反复地拨叫之后，仍然“铃声响，没人接”。格罗夫斯和卡汉（Groves & Kahn , 1979）指出，尽管对这种类型的电话号码试图进行的联系达 12 次之多，但最后可确定是合格住户的电话号码仍然屈指可数。

此外，并非所有的合格住户最后都是可以完成调查的。有些情形，如拒访或语言问题可能会将有些住户排除在调查之外。也有这样的可能，选到的被调查人因生病离开了城镇或因其他原因而无法接受调查。然而，回答率或在合格住户中完成调查的数目是必须估计的。无独有偶，这些数字最好也是从以前所做的调查中或试调查的结果中进行估计（基于试调查的估计可能有较高的抽样方差。其原因何在?）。

现在，我们再回到表达式 8.1，以确定最终完成 500 个调查所需选取的样本数。我们用列样本的结果作为我们分母中的第一项的估计值（1.00 - 0.089 = 0.91）。此外，我们假定我们发现在试调查中，有 28% 的 RDD 号码是

不合格的，且回答率是 70%。我们把这些数字代入表达式，得到如下的结果：

$$抽取的样本数=\frac{500}{0.91\times 0.72\times 0.70}=\frac{500}{0.458\ 64}=1\ 090$$

这就是说，我们必须从我们假设的电话簿中选取 1 090 行，才能最终完成 500 个调查。如果我们的假定是正确的，那么得到的结果将如下所示：

共有	1 090	从电话簿选取的行
其中	98(9%)	将是不合格的(即空行或营业性号码)
	992(91%)	是住宅电话号码

我们将这些电话号码的最后两位删除，代之以两位的随机数，生成新的电话号码。

在	992	个电话号码中，我们估计：
其中	278(28%)	将是不合格的(即空行或营业性号码)和不运行的
	714(72%)	将是合格的住户
在	714	合格住户中，我们进一步确定
其中	214(30%)	结果将会是调查不成功(即拒绝接受调查、状况不明和无法联系等)，这样最终将有
	500(70%)	个完成调查

选取样本

现在让我们来讨论从电话簿选取样本的问题。我们假设电话簿共有 440 页，每页 4 列，每列 104 行。我们希望选取一个系统的随机样本。首先，我们必须注意，在每一页进行的选取可能不止一次。实际上，我们在每一页需要进行 2～3 次的选取。我们究竟应该怎么在每一页进行选取呢？首先我们来比较一下总选取数和总列数：1 090/1 760 = 0.619。根据这一计算结果，我们应该选取 62% 列，或每 10 列选取 6 列。这样的间隔并不是很好处理的。我们希望选取的过程是简单易行而又随机的。注意 62% 与 66.7% 或 2/3 十分接近。一种可能是我们选取的样本比我们需要的更多，然后再从选取的结果中随机地删除多余的样本。这样剩余的样本仍然是随机样本，因为一个随机样本的随机样本仍然是一个随机样本。

最容易的做法是每 3 列取 2 列，即可能只需每 3 列删除 1 列。具体做法是先在 1 和 3 之间，选择一个随机数，然后不断将 3 加到这一随机数上，直至

和数超过 1 760。每一以这种方式选出的列,都将被排除在样本之外。根据我们选定的这一随机数,可能将有 587 列被删除,剩下的还有 1 173 列,我们将从它们之中选取样本。为了从这 1 173 列中选取样本,我们需要在 1 到 104(每一列的行数)之间选择一个随机数。用随机数码表,我们选择了 017 这一数。然后,我们在每一选出的列中,检查它的第 17 行。这时我们可以使用一块模板,测量每列的第一行到第 17 行的固定距离,把它们一一找出。如果找出的那行是居民户,我们便将它选入样本;如果是营业场所或空行,我们认为它是不合格的,便马上转移到下一个选取的列。

我们再一次来看一下表 8.2。假定第 3 列被我们从样本中删去,所以我们要检查的是第 1、2 和 4 列的第 17 行。第 1 列的第 17 行,列出了 A. Paul and Wanda Smith 的住宅电话号码,他们居住在 Ashe。我们选取他们的电话号码。我们在第 2 和第 4 列分别选取了 B. Steve Smith 和 D. Raymond Smith 的电话号码。对于所有被选取的电话号码,我们都要将它们的最后两位数删除,并代之以两个随机数。我们从随机数码表中,选取了数字 07,79 和 91。用这些随机数来取代我们选出的电话号码的后两位,产生了以下新的电话号码:451-6507,787-7879 和 751-0491。这就是我们必须仿照的、用以选取所有剩余的电话号码的程序。

如果我们的样本估计数是正确的,我们有望遇到 106 个空行及营业性电话和 1 067 个住宅电话号码。如果其他的假定也是正确的,我们还可预计,有 28% 的新电话号码是不合格的(n = 299),并最终得到 768 个左右合格的住宅电话号码。如果我们的回答率是 70%,我们最终将可能完成 538 个左右的调查。如果完成一个调查的费用为 30 美元左右,那么额外的 38 个调查需要增加的费用为 1 140 美元。这个数字也不是微不足道的小数。因此我们应当设法使我们最终完成调查的样本数接近 500。如何才能达到这一目标呢?我们将在图 8.1 对它进行具体的讨论。

一个极为重要的问题是,我们必须认识到,我们加到分母中的比例只不过是一个我们预计可能发生的估计数。随着选号和调查的进展,在任何一个阶段,实际的结果都有可能优于或差于我们的预期。我们最终得到的合格号码数可能少于估计数,但也可以得到一个比预期更好的回答率。这就使得我们得到的样本数多于我们所需要的,也可能得到比我们预期更多的不合格数和更低的回答率,这就使我们最终得到的样本数不足 500。因为任何一个组合数都有可能是最后得到的实际结果。所以对于研究者而言,重要的问题在于密切注视整个过程,以防患于未然,在必要的时候及时采取补救的措施。

下面的例子告诉我们,怎样才可以通过再次抽样(subsampling)和监控,使最终完成的调查数最大限度地与我们预期的数目相一致。图 8.1 展示了一个这样的例子。

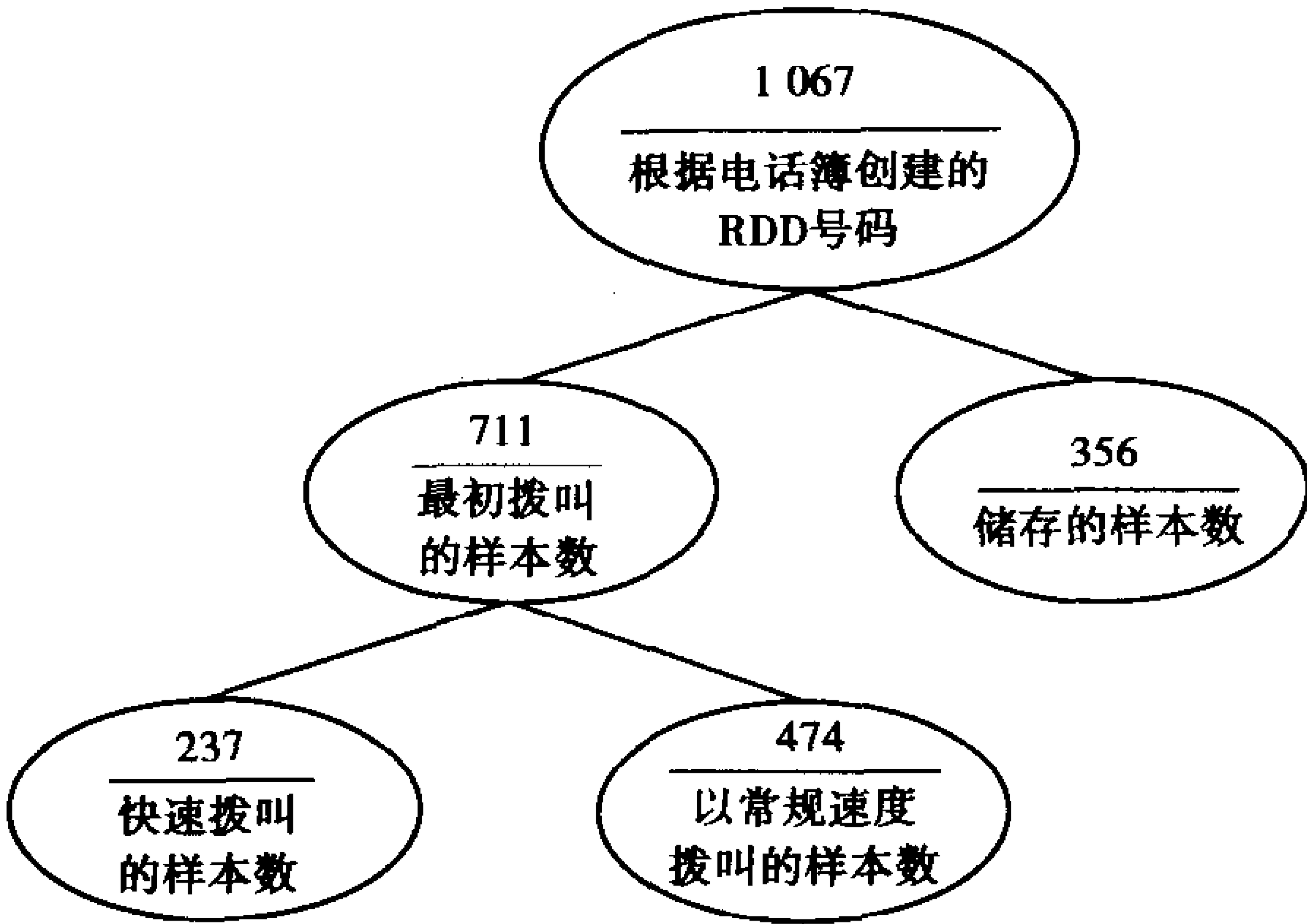


图 8.1　RDD 样本电话号码的处理方法

假定我们估计的空行和营业电话所占的比例是正确的，且样本产生的 RDD 号码是 1 067 个。每一个电话号码必须以随机和使偏倚为最小的方式分发给调查员。因为我们特定的电话号码样本有可能产生少量的不合格者和/或较高的回答率，我们并不希望一次就分发 1 067 个号码。我们希望对这些号码做二次抽样，看看这些二次抽样抽出的个案会有什么情况发生，进而用那些信息为下一个星期或下几个星期的调查活动做出安排。为了做到这一点，我们试图预期我们的电话号码能得到最佳的可能结果。我们把它看作一个样本的分布问题。这就是说，我们试图确定什么样的结果可能位于我们的合格电话号码比例的估计数 72% 和回答率的估计数 70% 正向的 3 个标准差。因为我们不知道这些信息，因此我们需要对之做一个有根据的猜测。我们用 85% 和 80% 分别做合格的电话号码的比例和回答率。我们再一次使用表达式 8.1，但不含空行和营业性电话号码，因为那一阶段的工作已经完成了。我们得到的结果是：

$$\text{样本选择数} = \frac{500}{0.85 \times 0.80} = 735$$

我们希望从 1 067 个样本号码中选取一个含 737 个号码的系统随机样本。为了能抽取这样一个样本，我们可以参照前面从电话簿中选取列时的做法。首先，我们求得 735/1 067 ＝0.689，它与 0.667 或 2/3 十分接近。尽管 0.667 × 1 067 ＝ 711，比原数少了 24 个，但是我们从这一样本数开始我们的调查则是合乎情理的。考虑实际的可行性，将选取的 711 个号码分为两个子样本似乎不失为一种明智之举。这样我们可将其中一个用于快速调查，以检测我们的假定，进而确定是否还需要额外的样本号码。对另一个样本的调查可同时进行，但是速度则按预先规定的标准控制。快速调查的样本的样本量可以是为 711 个号码的三分之一，即一个含 237 个电话号码的子样本。随着

数据收集工作的进展，我们继续对照我们的假定，对调查结果进行监控。与此同时，我们手中还握有一个储备样本，它是一个含 356 个号码（1 067 - 711 = 356）的随机的子样本。从这一储备样本中，我们还可以抽取若干随机子样本。这些子样本的电话号码可按需要分期分批地发放给调查员。如果需要，也可一次性地全部发放给调查员。

例三：其他 RDD 电话号码样本

还有其他若干种可以用于选取 RDD 样本的方法。其中一种是先确定地区电话的前置号即地区号，然后在它们后面加上一个四位的随机数。与前两个例子相比，这一方法使得抽样工作变得比较容易，但是调查却会变得更加困难。这一点我们很快就会看到。

表 8.3 罗利和选取的附近城市的前置号（地区号）

地点	前置号码
阿佩克斯（Apex）	303，362，387
卡里（Cary）	319，380，460，467，469，481，677
加纳（Garner）	662，772，779
纳尼特戴尔（Knightdale）	217，266
莫里斯维尔（Morrisville）	467，469，481
罗利（Raleigh）	212，231，233，250，301，302，420，501，505，508，512，515，516，518，546，571，662，664，676，713，715，733，737，740，755，772，779，781，782，783，787，790，801，821，828，829，831，832，833，834，836，839，840，846，847，848，850，851，856，859，860，870，872，876，878，880，881，890，899，954，971，976，981
温德尔（Wendell）	266，365
泽布龙（Zebulon）	269，365

电话簿前面部分刊登的一般都是在当地拨叫时的住宅电话的前置号码。表 8.3 列出了在一本每年校刊一次的电话簿上，刊登的罗利和选取的附近城市的前置号。我们称之为早期关注区（area of earlier interest）。不难发现，前置号的数目与城市的大小有关。罗利是其中最大的城市，因此占有的前置号也最多。其次是卡里和加纳。此外，多数城市，但并非全部城市都有独特的前置号。例如阿佩克斯就拥有自己独特的前置号，而莫里斯维尔却是卡里的前

置号集合中的一个子集。纳尼特戴尔、温德尔和泽布龙三地的前置号则有一定程度的重叠。独特的前置号使研究者得以专注于某一特定的地区,如果这是他们所希望的。如果前置号是重叠的,而研究者又只是对某一城市感兴趣,那么在调查开始的时候,我们就必须使用筛选性问题,对被调查人居住的确切的地理位置加以确认,去除不合格的被调查人。

在这一地理区域共有 77 个独特的前置号。为了生成 RDD 电话号码,我们只要在前置号后面加上四位的随机数字就可以了。但是,我们究竟需要多少个号码呢? 我们可以使用表达式 8.1 中的假定。如果我们希望做 500 个调查,那么公式中的分子便与原来的相同。因为我们并非从一本电话簿中选取电话号码,所以我们不必处理空行和营业性电话。此外,我们还要去掉电话簿中住宅电话的百分比,因为它不适用于目前的情况。

我们必须估计不合格数和回答率。我们假定回答率与早先一样为 70%,这样表达式中的未知数就只剩下一个。我们是否可以再一次用 72% 作为合格的电话号码的比例呢? 答案是否定的。因为我们用以生成电话号码的方法已大不相同。主要的差别在于我们现在使用的方法是在前置号后面加上四位的随机数字,而非从现存的电话号码中去掉最后两位。从表面上看,二者没有什么很大区别,但是实际上并非如此,我们很快就会给读者阐明,二者可能的差别是什么。

在前置号之后的四位号码叫做后置号。每一个前置号后面附有的后置号可能有 10 000 个之多,它是从 0 000 ~ 9 999 之间的任意一个四位数。在我们从一本电话簿中选取一个电话号码,并去掉其中最后两位数时,我们已经知道电话公司已经分配给一个后置号的前两位数。早先,我们选取了电话号码 787-7831,我们知道电话公司用于前置号 787 的后置号是 7_ _ _和 78_ _。后者就是一个百位号码库。在将四位数字加到前置号后面时,在拨叫这一号码之前,我们一般都不知道电话公司使用的后置号。因此,使用四位数的随机后置号会导致更多的未运行的电话号码。

在将四位的随机数加到前置号后面时,我们可以用普查数据和了解掌握独特的前置号数来估计不合格号码所占的比例。我们假定关注的罗利地区,在调查前 3 年,实有住户 144 621 户。我们用普查数据对它做了验证,并发现这些住户的 96.5% 或 139 559 户内有电话。因为我们使用的普查数据是大约 3 年前的数据,因此我们需要对这一地区的住户的增长情况做一个估计。一种估计这方面情况的方法是以前 10 年的增长数字为参考,并假定这一地区的增长率不变。另一种方法是从房地产公司、银行或政府机构得到估计数。在查阅了普查数据之后,我们发现在前 10 年期间,威克县的城镇人口增长了 49%,但农村地区只增长了 19%。以过去 10 年的这一数字为依据,我们假定这一地区的增长率为 14%,有电话的住户数为 159 098。表 8.3 显示,该地区

共有独特的前置号 77 个。因为每一前置号都可以有 10 000 个可能的独特的后置号，所以这一地区总共可有 77 × 10 000 = 770 000 个独特的电话号码。合格电话号码数的比例的估计值是 159 098/770 000 = 0.206 6。我们需要创建的四位随机数字数取决于我们在表达式 8.1 使用的新假定[3]。

$$\text{抽取的样本数} = \frac{500}{0.2066 \times 0.70} = \frac{500}{0.1446} = 3\,457$$

这一结果比我们在例一中以电话簿为根据的方法求出的那个需要的电话号码数 992 大得多。这说明，我们大概需要拨打 7 个电话，才能完成 1 个调查。我们拨打的许多电话和一大笔的预算将主要用来去除那些未运行的号码。

现在我们再来继续讲解我们的例子。我们共有前置号 77 个，且估计为了完成 500 个调查，我们共需要电话号码 3 457 个。我们需要为每一个前置号建立的四位随机数是 3 457/77 = 44.89 或 44.9。四舍五入取整后为 45，使总数增加了 8，差别并不大：45 × 77 = 3 465。因此我们需要为每一前置号选取 45 个四位的随机数。我们可以从一张随机数码表中选数，也可以用计算机生成四位的随机数。因为我们的总数 3 465 是一个估计数，所以我们需要构建所有的电话号码，然后再从这些号码中取一个系统的随机子样本，仿照我们例二所做的和图 8.1 所介绍的那样，拨叫子样本中的每一个电话号码，并监控其结果。

在这一例子中，调查员将要用相当多的时间来去除未运行号码。然而，我们可以找到一些办法来帮助我们减少未运行号码的数目。海尼斯和科公司（位于俄亥俄州海特市的中堡）为选中的城市出版了一种十字形的电话簿。这种电话簿以前置号和后置号的大小，而不是以客户姓氏笔画，由小到大来排列客户的地址。这种排列法便于我们找出未在使用，并可以在选取过程中去除的后置号。例如，假定两个最小的前置号 212 和 217。因为后置号是以升序排列的，我们也许会注意到，在那些有最小的前置号 212 的电话号码中，最后一个电话号码是 212-4 356，而紧挨着它后面电话号码是 217-0 018。那么我们可以据此认为在前置号 212 中，后置号 5___、6___、7___、8___和 9___尚未使用。如果情况的确如此，那么我们就能从我们的可能的电话号码清单中排除 5 000 个未运行号码。我们可以对整个清单做类似的检查。

在那些无法得到十字形电话簿的地区，有时我们可以审视电话簿页码的大样本，试着将它们以前置号和后置号的大小按升序排列。但这一工作必须慎之又慎，因为一个后置号一旦被排除，那些使用这一号码的被调查人将没有机会进入我们的调查样本。最后，大的号码段可以通过检查这一地区的大雇主是否给自己的公司配置了特有的前置号来加以删除。例如在罗利，北卡罗来那州立大学使用前置号 515 和 513，北卡罗来那州的前置号是 715 和 733。

我们可以很有把握地删除 40 000 个可能的电话号码。某些未运行电话号码和营业性电话在采取所有这些措施之后，可能仍然没有被去除掉。营业性电话号码最好一一加以确认，并在做第一次试拨的当天，即星期一到星期五之间的某一天再去除。

全国性 RDD 样本

从 20 世纪 90 年代开始，电话调查系统在全世界和美国经历了重大的变化。新的技术成为它的主要推动力，而从那时开始的变革在进入 21 世纪之后，仍然在继续。人们都在猜想，2010 年，这一系统将会是什么样的呢？已经在美国的系统中发生的某些重大变化是号码、分布和百位号码库的住宅电话的密度。塔克等人(Tucker et al. ,2002)在一份报告中说，含一个登录的住宅号码的百位号码库的数目在 1990 年是 38%，到 1999 年下降到了 30%。在 1990 年，435 万个百位号码库的 62% 没有任何一个登录的住宅号码，到了 1999 年，这一数字上升到了 772 万个的 70%。就电话的密度而言，在 1990 年，所有的百位号码库的 14.9% 有 59 个或 50 个以上的电话号码，而在 1999 年这一数字是 3.5%。这些变化对许多设计功效都会发生明显的影响。

前面已经讨论过的许多方法也可以用于全国性 RDD 电话调查样本的选取。差别在于全国性电话调查样本的选取，需要使用全国性的样本框，且某些设计的电话号码的选取，需分两个阶段进行。从一些电话簿中选取全国性的样本虽然是可能的，但是效率却不是很高。美国每年出版的电话簿有 5 100 本之多。这些电话簿不仅厚薄不一，而且每年出版的时间也不尽相同。购买这些电话簿，并用它们选取样本不仅费用十分昂贵，而且旷日持久。所幸的是，我们可以找到更加好的方法。

在过去的十余年间，清单辅助的 srs 抽样方法便是可供我们选择的一种不错的抽样方法，这种设计和选取方法，我们已经在例一和例二中做了介绍。另一种广为人们使用，且出现早于清单辅助的 srs 抽样法是梅托富斯基和沃克斯博格(Waksberg, 1978)提出的。这种方法使用由卓讯科技公司(TT)生产的数据磁带。TT 公司生产一种按季度更新的清单磁带，叫做运行代码清单(NPA/NXX Active Code List)，它收录了美国每一地区的地区码和正在运行的前置号。在 2003 年中，清单登录了 271 个地区的代码，且有地区码和前置号的组合 86 893 个。我们可以用 TT 数据磁带作为全国性的 RDD 调查的抽样框。先选取一个地区码和前置号组合的样本，然后再加上四位的随机后置号。不过这样选取的后置号会遇到例三曾经遇到过的同样的问题：有许多未运行的号码。用这种方法，塔克等人(Tucker et al. ,2002)发现只有不到 15% 的样本号码才是正在运行的住宅号码。为了解决这一问题，梅托富斯基和沃克斯博格(Waksberg, 1978)建议使用一种方法，这种方法虽然能使拨号的效率有

一定程度的提高,但仍不如 srs 清单辅助的抽样设计的效率高。样本选择和调查是分两个阶段进行的。最初,研究者必须确定需要调查的总数及第一和第二阶段需要选取的抽样单位数。让我们来做一个实例,因为它不仅可以告诉我们怎样做多阶抽样,而且也可以阐明概率与样本大小(某一量度的)成比例的抽样法的原理。

假定我们想要在 400 个地区完成 1 600 个调查(我们在后面对地区进行定义),我们用从 TT 数据磁带选取的地区码和前置号组合的系统随机样本,生成用于第一阶段的、在全部 400 个地区进行的调查电话号码。我们需要选取多少个这样的地区码和前置号的组合呢?为了确定这一个数字,我们将 400 这一数除以被塔克等人发现的合格号码比例(Tucker et al. ,2002),得到的结果是:

$$\text{第一阶段的选样数} = \frac{400}{0.15} = 2\ 667$$

因此我们决定先从 TT 全部 86 893 个地区码和前置号的组合中选取一个容量为 2 667 的系统随机样本[4],然后再将四位随机数加到六位的地区码和前置号的组合后面生成电话号码。这 2 667 个号码最后终将产生 400 个住宅电话号码。不过,我们知道并非对所有这些号码的调查最终都能完成。我们已经估计到回答率可能在 70%,因此我们可以预期,最后完成的调查在 280 个左右。

我们将样本一分为二,分别用于抽样的两个阶段,因为我们从前面的例子已经知道,一旦我们找到了一个住宅号码,这就意味着我们有可能在这一地区码和前置号下找到更多的住宅号码。梅托富斯基和沃克斯博格设计的第二个阶段是保留已在第一阶段被确定为住宅电话的电话号码的前 8 位,及地区码、前置号和四位后置号的前两位。我们不断地用一个又一个的二位随机数,加到每一个这样的 8 位数之后,直至我们通过电话联系确定,经这样的修正之后的确产生了一个正在使用的 8 位数的住宅电话号码为止。例如,假定号码 919-387-7892 是一个住宅电话号码。为了生成第二阶段的电话号码,我们取出了最后两位,并不断地代之以其他的两位随机数,直至我们确定修正后的号码的确是又一个住宅号码为止。

为了将我们的例子做完,我们需要生成多少个其他的号码呢?在第一阶段,2 667 个号码生成了 400 个住户,并最终完成了 280 个调查。因此我们还需用 400 个第一阶段的电话号码完成 1 320(1 600 - 280)个以上的调查。这就是说每一个第一阶段确定的住宅号码,平均需完成 3.3(1 320/400)个调查[5]。那么我们究竟需要选取多少个第二阶段的号码呢?塔克等人(Tucker et al. ,2002)发现,49% 的第二阶段号码是住宅号码。如果我们仍然用 70% 作为回答率,表达式可以用下面的方式求解:

$$\text{第二阶段平均选样数} = \frac{3.3}{(0.49 \times 0.70)} = 9.62$$

每一个第一阶段的住户号平均生成9.62个电话号码将导致:为了最终完成1 320个调查,我们需要拨叫3 848(9.62 × 400)多个号码。第二阶段的号码的生成方法是去掉第一阶段确定的住宅号码的最后两位,然后不断地代之以一组独特的两位随机号,直至修正后的号码可以确定是住宅号。

梅托富斯基—沃克斯博格设计是一种二阶整群样本,在第二阶段之后,它具有相等的选择概率。第一阶段的100个号码组的选择概率与百位号码库拥有的住宅号码数成比例。因此,百位号码库中的住宅号码越多,选择的概率就越高。正因为如此,为了使整个的选择概率相等,在第二阶段选择的住宅单位数必须是一个既定的数。

我们可以把这些原理说明如下。假定电话号码A是一组电话号码,它共有电话号码100个,其中80个为住宅号码。B则是另一组电话号码,总数也是100,但只有住宅号码20个。如果从两组中各选取一个电话号码,那么组A入选样本的概率就是80/100 = 0.80,而组B则是20/100 = 0.20,在第一阶段,组A入选样本的概率是组B的4倍。在第二阶段的选取时,我们希望能再确定4个住宅电话号码,以使每组住宅号码总数都是5个。每组每一号码选择的总概率都等于这两个概率的积。诚如我们在下面所示,在第二阶段之后,选择的概率就相等了[6]。

$$\text{组 A} = \frac{80}{100} \times \frac{5}{80} = 0.8 \times 0.062\,5 = 0.05$$

$$\text{组 B} = \frac{20}{100} \times \frac{5}{20} = 0.2 \times 0.250\,0 = 0.05$$

这一种设计方式引进了一些复杂的因素,它使选取的样本不再是一个简单随机样本。首先,用于计算样本方差的公式必须考虑到二阶设计问题(Kalton, 1983)。其次,因为住在同一地区的人们的态度和行为,较之整个总体有更多相似之处。而这种类型的样本具有聚类效应,因为聚类样本内的人们往往比总体的随机样本具有更多的同质性,因而会低估总体的方差。这将导致某种设计效应(design effect)——在样本量相等时,聚类样本的方差与总体简单随机样本的方差的比率。第三种复杂性是样本的处理必须连续进行。不仅如此,处理过程还需要辅之以大量的监控,以确保确定的号码不超出类(百位号码库)的范围。最后,在有些情况下,确定的电话号码数达不到类别规定的数目。严格地讲,在这样的情况发生的时候,我们需要进行加权。权数是住宅电话号码的目标数和实际确定的电话号码数的比率。既然梅托富斯基—沃克斯博格抽样法如此复杂,那么我们为什么还要使用它呢?原因在于聚类样本的抽样和数据收集的费用大大低于简单随机样本,因此在设计效应

相对那些关键性的变量比较小的时候，人们都乐于采用这种抽样法。

怎样在住户内选取被调查人

一个电话号码的样本是一个住户的样本。这样的样本有可能收集住户层次的有关活动相当精确的信息，例如过去两周内，户内所有人员看医生的次数；过去12个月内住院的天数；过去6个月内购买的耐用消费品的数量和种类；过去12个月内拥有的轿车数和行驶的总里程数等等。在做诸如这样的调查的时候，我们拨叫一个电话号码，并告诉接电话的人我们想要调查一些什么。我们希望能与户中最了解有关活动情况的人交谈。这种类型的调查试图对能够提供最为精确的信息的成年人进行调查。

在许多情况下，我们希望把我们的调查结果推广到整个总体，或对总体的某些特征加以概括。如果我们调查的是接电话的那个人，那么在调查结果中，青少年、妇女、那些在社区事务中比较活跃的人和那些在家中经营者的比例将会过高。为了满足我们对总体进行概括或对个体的某些特征加以估计的需要，我们一般需要在住户中随机地选择一个人，并以户内合格的被调查人数对调查数据进行加权。加权是必须的，因为住户是合格个体的聚类。在每一户选取一个个体时，那些生活在户内有两个或更多个合格的被调查人的人们选择概率，与那些生活在只有一个合格的被调查人的户中的人并不相等。一人户中的人有选择概率1.0或100%。因为该个体是户中唯一合格的被调查人。二人户中人的选择概率是0.5，而3人户中人的选择概率则是0.33，依此类推。

经典的户内选择法是凯思（Kish，1965）为面对面调查设计的，并已经被应用于电话调查。在电话调查中，在做了一番简短的介绍之后，调查员要求被调查人分性别和年龄报出户内所有的合格的成员。然后调查员必须给每一个合格的成员，从年龄最大的男性成员开始，依次编号，编完了男性，再接着编女性，也从年龄最大的开始，依次编号，直至编完。使用事先选定的随机数——二人户的数字是1或2——调查员选取那个编号与预先预定的随机数对应的个体作为被调查人。凯思的方法虽然是无偏倚的，但是它需要在调查开始的时候耗费大量的时间。在对调查的合法性和信任尚未建立之前，许多被调查人都会对如此详细的列举户内成员的问题心存疑虑。正因为如此，许多调查研究者认为这一选样方法可能会导致较高的拒访率。

还有一些其他的方法，它们是特洛汉和卡特（Troldahl & Carter，1964）及布赖恩特（Bryant，1975）提出的。这些方法旨在以较少的唐突进行户内人员的列举。在使用这种方法时，接电话的成年人被要求回答两个问题：

1. 包括您自己在内,您户中共有多少个18岁或以上的人?

2. 其中多少个是男人(女人)?

四种或七种选择矩阵中的某一种被随机地分配给每一个样本号码[7]。选择矩阵并不要求把成年人列表。而代之以由调查员根据住户的人员组成情况,要求与户中"最老的"或"最年轻"的男人或女人交谈,甚至更为简单地要求和一个男人或女人交谈。在一个对这些方法进行比较的实验中,我们并没有发现特洛汉—卡特—布赖恩特法在某些方面优于凯思法。无疑,我们首选的方法应该还是凯思法。

更新的户内抽样法是由奥洛克和布莱尔(O'Rourke & Blair,1983)及沙蒙和尼克尔斯(Salmon & Nichols,1983)提出的。这些方法不需要列举住户成员。调查员只要求和户内合格成员中"最近"或"最快"过生日的人交谈。最近生日法的操作方法如下所示:

您好,我从马里兰大学给您打电话。我们正在做一个研究,了解一下华盛顿大都市地区的人们对那些与自己有关系的事情的看法。您的家庭被随机地抽到参加我们的调查。这次调查,我们需要和您家中年龄在18岁或以上的、最近过生日的成年人谈一谈。您家中谁适合这样的条件?

如果这个人不知道所有人的生日,那么可接着问:

在那些您知道他们的生日的人中,谁是最近一个过生日的人?我们想和他谈一谈可以吗?

两种方法似乎都给了每一个合格的成员相等的选择的机会。最近生日法可能比下一生日法的选择误差要小一些,因为回忆一个过去的事件比了解一个将要发生的事件要容易一些,除非将来的事件很快就要发生。我们需要进行方法论意义上的研究,以确定这些假设是否真实,并更好地估计这些选择方法的信度和效度。奥丁迪克、比绍普、绍瑞生和塔克法比尔(Oldendick, Bishop, Sorenson, & Tuchfarber,1988)曾经做过一个方法论上的研究,对凯思法和最近生日法做了比较。这些研究者们对拒访率、样本的人口学特征和大量的问题的答案进行了比较。他们发现具有统计意义上的差异可谓屈指可数,于是他们认为两种方法并不存在实质性的差异。我们仍然需要做更多的诸如这样的研究。

最近,瑞佐、布瑞克和帕克(Rizzo, Brick, & Park,2004)提出了一种新的方法。这种方法利用了事实上只有一或两个成人的住户占到了美国全部住户的85%左右这一有利条件。在与一个住户进行联系的时候,我们提出的第一个问题是:您家中有多少个年龄在18岁或18岁以上的成年人?在遇有回答只有一个的情况时,这个人无疑就是我们选定的被调查人。在遇有户中有两个成年人的时候,一半时间取接电话的人做被调查人,另一半时间取另一

个成年人做被调查人。只有15%左右的住户需要使用比较复杂的方法，在遇有这样的情况时，我们既可以用“最近生日”法，也可以用上面使用过的其他方法。

例四：学生名单样本

教授或其他学生经常会进行一些有关大学生的调查。我们已经讨论过的那种用于这样的调查方法是一种电话调查[8]。而在这一例子中，假定我们只有很少一点的资助，但时间却不成问题。因此我们决定尽可能多地使用校园邮件系统，做一个邮寄式调查[在例五中，我们将讨论另一种可能性——为一个集体代理调查(group-proctored survey)选取班级的样本]。

选取抽样框

注册办公室有最近的注册资料。在调查有很严肃的学术关怀和收集的资料对学校的其他团体也很有用处的时候，这些部门的态度是很合作的。几乎所有的大专院校都将学生的资料保存在计算机文件中。这些资料通常都以姓氏笔画为序。我们假定我们只想调查本科生，而我们的大学有本科生20 000个，其中一年级新生8 000个，二年级、三年级和四年级学生各6 000，3 000和3 000个。我们想要选取1 000个学生。一种方法是要求注册办公室从所有的本科生的计算机文件中选取一个1 000个学生的简单随机样本。尽管我们不能保证样本学生的年级比例将与总体相同，但对于一个这样的容量的样本，其中应该大约会有一年级学生400个、二年级学生300个、三年级学生300个和四年级学生150个。

确定抽样方法

如果样本成员的主要特征(入校年限、性别)可能与重要的因变量有关，我们可以用一种比较好的抽样方法。这种方法叫做分层抽样。苏德曼(Sudman，1976)提出了四个对样本进行分层的理由：

1. 群体是研究的兴趣目的所在。
2. 方差因群体而异。
3. 调查费用随群体的变化而变化。
4. 前期的信息量因群体而异。

在分层抽样中，我们希望将总体分成由类似的个体组成的群体。我们希望群体内的元素是同质的，而群体间则是彼此有差别或异质的。作为分群标准的变量或一些变量必须与因变量有关，否则的话，我们为分层而做的一系列

工作将会是徒劳无益的。

使用隐含的分层的意义在于它是一种通过确保样本在特定的变量上的分布比例来提高抽样的效率的方法。让我们再回到例子,假定我们确信我们的关键性因变量将随年级的变化而变化。这样,我们将要求注册办公室把学生按年级重新排序列表,而不是从一个按姓氏笔画排序的名单上选取一个简单随机样本。前 8 000 个学生将是一年级本科生,随后则是 6 000 个二年级的本科生,随后是三年级和四年级的学生。然后我们再从这一份名单上选取一个系统随机样本。因为总体的容量是 20 000,而样本的容量是 1 000,所以我们的抽样间距是 20。我们在 1 ~20 选取一个随机的起始点。这样,我们得到的样本将与总体成比例。以这样的方式使用隐含的分层,我们可以确保样本将包含 400 个左右的一年级学生、300 个左右的二年级学生、150 个左右的三年级学生和 150 个左右的四年级学生。

我们可能有一份根据多个变量排序的名单。例如,我们确信性别也是一个重要的变量,那么我们就可以将名单同时按年级和性别排序:一年级的男生可以先列在名单上,随后是一年级的女生,随后是二年级的男生,接着是二年级的女生,等等,依此类推。选取一个系统随机样本可保证性别和年级的分布比例有代表性。

我们在第 7 章中曾经说过,用于群间比较的最优设计是每群的样本量相等的设计。在我们的这一个例子中,不同年级的学生的样本量是不等的。为了在校内的每一个年级选取等量的样本,我们可以把学生按年级分成 4 组。在每一组内,我们都选取一个容量为 250 的系统随机样本。用于一、二年级的抽样间距为 1∶32 和 1∶24,而三、四年级都为 1∶12。这样的样本叫做不成比例的分层样本(disproportionate stratified sample)。虽然每一组的样本量和完成的调查数是相等的,但是抽样比和不同年级的学生的选择概率却是不相等的。在分析我们的数据的时候,可以分年级报告我们的调查结果,例如一年级新生上周饮酒三次以上的如何如何,二年学生上周饮酒三次以上的如何如何,等等。但是我们不可简单地将这四组数据合并在一起,因为它们的选择概率是不同的。如果我们希望把数据合并在一起,那么就需要对数据进行加权。对于所有的学生而言,相等的选择概率是 1∶20。因为一、二年级的学生被抽少了,必须给他们以大于 1 的权,而三、四的学生都被抽多了,所以应该他们一个小于 1 的权。以这种方式对数据加权,使我们得以将来自不同的组的所有信息合并在一起,对在校的所有的本科生上周饮酒三次以上的比例做一个估计。

例五:大学班级样本

一个十分高效的选取大学生样本的方法是选取一个班级的样本。这种方法利用了自然的聚类和被调查人受制于纪律不能缺席这两个有利条件。如果班级学生的平均数是24,那么原本只可以完成1个调查的时间,现在可以多完成23个。这种方法的关键是得到被选出班级的教授的合作。

定义总体和准备抽样框

和一般的抽样方法一样,这种方法也从定义总体开始,例如将总体定义为在校的本科生和得到抽样框,如目前在校的班级的清单。我们必须保证将那些有可能被排除在外的班级包含在抽样框内,而将那些已经取消的班级排除在抽样框外。不仅如此,我们也必须将重复出现的班级清除掉。我们必须对跨系科的大课班和附属的试验课班或讨论课班加以定义。如果一个大课班有许多学生,譬如说100个,而试验课的人数却比较接近班级的平均人数,那么我们便可以将实验课班保留在样本框内,而将大课班排除在外。其原因是,大课班的学生是高度跨班相关的,势必增加估计的方差。

在北卡罗来那州立大学的课程表中,班级是分系按班级名称的字母顺序排列的。在系内,课程则按程度由低到高排列的,既从导论课到最高级的专业课。我们用系统随机抽样等概地选取班级的样本。按字母顺序排列的清单不利于做分层抽样。如果我们认为态度和行为是因学院而异的,那么我们可以要求注册办公室生成一个新的、分学院排列的,且在每一学院内,按课程的程度由低到高排列的班级清单。

求样本选取数

在选取班级之前,我们需要确定框内班级学生的平均数、本科生平均选修的课程数、预期的教授和学生的合作率和要求最终完成调查的学生数。班级学生平均数和学生选修课程的平均数,可能可以从注册办公室拿到。只需要做一些大致的估计。最重要的信息,每班学生的平均数如果不知道,可以用注册学生数乘以平均选修课程数,再除以有关部门提供的班级数来估计。用这种方法我们求得:

$$\text{班级平均人数} = \frac{20\ 000 \times 4}{3\ 320} = 24.1$$

假定我们希望完成1 200个调查。如果每班平均有学生24个左右,那么我们就需要调查50个班级。现在我们已经知道,并非每一个教授都同意我们

到他们的班上进行问卷调查。因此我们还需要估计合作率。无独有偶，我们再一次去看一看以往的调查或做一些试调查。在马里兰大学曾经做过的类似调查中，公园学院校区 65% 的教授配合了调查，而在选定的班级中，99% 的学生都配合接受了调查。如果我们使用 65% 作为合作率，那么我们就需要选取 77 个班级进行系统随机抽样。

维护设计的整体性

虽然这种设计是很有效的，也是很经济的，但是为了保持设计的整体性，研究者还必须再考虑三个问题。第一，这种方法将产生一个所有注册学生的概率样本。但是，学生的选择概率是不等的，因为选修课程较多的学生较选修较少的学生，有更高的入选机会。正因为如此，恰当的数据分析是给每一个学生一个权。这些权可相互修正，这样加权样本的总容量和未加权样本的总容量是相等的。因为平均选修课程数为 4，我们给那些选修 4 门课的学生的权是 1，给那些选修 4 门以上课程的学生以小于 1 的权，而给那些选修不到 4 门课程的学生以大于 1 的权。每个学生选修课程数则必须在问卷中询问。删除了这一问题将使我们无法得到对数据进行适当的加权所需要的信息。

第二，我们都知道，因为各种原因，所有的学生不会每一堂课都参加。如果在听课和缺席的学生之间，在因变量上存在着差别，那么如何将那些缺席的学生包含在调查的估计值中便会是非常关键的，否则的话，最后的结果可能会存在某种偏倚。对那些缺席的学生进行补充调查所需的费用，远比课堂上分发、回收问卷的要高。研究者既可以用邮件，也可以用电话对缺席者进行补充调查，具体应该怎么做，则要视时间表和预算而定。尼曼(Neyman，1934)设计了一种方法。研究者用这种方法计算补充调查的费用和课堂调查的费用，并以这两种费用之比的平方根决定需要做补充调查的缺席者的百分比。在数据分析的时候，对缺席者的子样本进行加权，以使分析结果反映所有的缺席者，并进而使最终的样本估计值是无偏倚的。

第三，一个聚类的随机样本不是一个最小单元的随机样本。我们希望聚类样本的类是由异质的单元组成的，这样一个聚类样本就能被看作是一个反映总体的样本。然而，自然的聚类一般反映的是总体的某些相似性，而不是多样性。实际上，登记选修社会学导论的学生更可能是一或二年级的本科生。而由这样的学生组成的样本并不是一个全体注册学生的随机样本。聚类样本的构成比总体的随机样本的构成有更多的相似性。而这种在构成上的相似程度就是我们所谓的聚类样本的设计效应。这一设计效应导致聚类样本的精确性比等量的 srs 样本小[9]。低效的程度取决于聚类样本的方差与等量的 srs 样本的方差之比。这一计算，超出了本书介绍的范围，那些对这一问题有兴趣的

学生可以查阅卡尔登的著作(Kalton,1983)。

注 释

1 某些电话簿上的住宅号码清单的计算机文件,可以从商业机构那里免费得到。

2 在第7章我们已经阐明,在所有其他条件相等的情况下,样本量越大,我们在样本估计值中可以具有的置信度就越高。一个样本量为25的样本是非常小的样本,因而它的抽样误差是非常大的。读者必须牢记,在将小样本用于预期目的时,最后的结果可能与预期的样本显著不同。

3 表示空行和营业性电话的百分比项已经从表达式中删除,因为它不适用于这种生成RDD号码的方法。

4 这是在2003年10月的组合数,有趣的是,1993年11月,组合数为44 129。

5 我们用400这一数字,因为它是在第一阶段确定的住户数。这是一个重要的数字,但不是最终完成的调查数。

6 这一例子已经被简化,因而在技术上已经并非总的选择概率。在第一阶段的抽样分数必须乘以概率0.05。但是最终的总选择概率必须相等。

7 选择矩阵5,6和7给男性以更高的选择概率,因为男性的合作率低于女性。

8 一种可能的抽样框是出版发行的大学学生电话簿。我们可以用基于电话簿的RDD样本的同样的方式,选取一个姓名的样本。不过在这一例子中,我们并不想产生随机数字电话号码。

9 在样本量相等的情况下,聚类样本的置信区间或样本标准误差大于srs样本。因此,它的变量总体值的估计值的精确性比较低。

第9章 减少数据收集中的误差

Reducing Sources of Error in Data Collection

本章要讨论的问题是有关数据收集过程中的不同方面存在的问题。这些问题都会对我们得到的结果的精确性有一定影响。要讨论的问题主要有两个:误差的概念和减少误差的方法。我们先介绍调查误差的一般概念,然后讨论偏倚和方差的概念,并对单位和题项的无回答问题做一个概括的介绍。在此之后,我们将对以这些概念为基础构建的调查质量的主要量度进行一些讨论。近年来单位无回答有所上升,我们提出了一个解决这一问题的理论框架,以便读者能据此精心挑选一些用于确定和解决这些问题的方法。

这些背景知识为本章第二部分内容的介绍提供了扎实的基础。第二部分的主要内容是介绍一些减少误差来源的方法[1]。某些误差来源具有普遍性,既存在于调查员填写的调查,也存在于被调查人自行填写的调查。我们首先讨论调查员填写的调查中的误差源,并指出那些同时存在于两种调查方式中的问题。最后,我们将专门讨论那些存在于自行填写式调查,如常规的邮寄式调查或互联网数据收集中的问题。

误差的起因

设想一下一个完美的样本调查:调查设计和问卷满足了所有的研究目标;我们可以得到的抽样框包括了每一个总体成员的精确信息;选取的样本确切地反映了总体所有的事实;测量工具中的每一个问题绝对清楚,并抓住了我们所感兴趣的问题的每一个维度;我们为样本选取的每一个人都联系到了,并立即同意参与我们的调查;调查员所做的调查完美无缺,且从不以他们自己的行为,甚至仅仅因为他们的存在而对被调查人的答案有所影响;被调查人能按研究者要求的那样确切地理解每一个问题,了解所有我们所需要的信息,且总是如实而全面地回答每一个问题;被调查人的回答被如实地做了记录,且没有任何误差地被输入了计算机文件;最终得到的那一套数据的信度和效度堪称

典范。

除非是一个微不足道的小例子，否则我们无法找到如此完美无缺的调查。做一个调查的每一步都存在令我们背离这样的理想状态的可能，有时这种可能很小，但有时这种可能则很大。正像调查过程中的每一步和每一个参与者都会为得到精确的信息做出贡献一样，每一步和每一个参与者也同样可能会使信息的精确性有所降低。我们把这种降低信息的精确性的可能叫做调查的误差源[2]。每一个调查都有调查误差，大多数误差都无法在我们有限的资源范围内完全被清除，有些误差甚至从样本调查的原理角度看，也是无法清楚的。

认识到试图做到完美无缺是不现实的这一点，使我们很快就会考虑到某些比较实用的问题：

- 我们计划进行的调查的误差源是什么？
- 在这些误差源中，哪些是我们最应该重视的？
- 我们可以采取一些什么样的行之有效的措施来减少这些主误差源？

诚如调查设计和实施过程中的所有方面一样，有关如何处理误差源的决策，也必须在一个对照的框架下考虑：设想如果未对一个误差源做出处理可能会对调查的精度造成什么样的危害？在明确回答这一问题之后，再权衡费用和其他的资源，然后方可做出决策。

有些误差源可能比其他误差源危害性更大。在我们能对一个特定的调查的误差源进行有效的鉴别之前，我们需要在普遍的意义上对误差源有所了解。这一了解有赖于两个概念：方差和偏倚。这两个误差源一起被用于对调查总误差大小的评估。就本书的目的而言，我们只要求读者能在一般意义上对这些概念有所了解就可以了，所以我们将主要通过一些例子来介绍这些概念。

方差和偏倚

方差是在用某种方法反复进行试验时出现的测量差异。尽管在大多数调查中，我们并不反复进行试验，但这是一个很有用的概念。我们已经对方差最为常见的例子——抽样方差做过介绍。大家不妨回想一下，如果我们选取了一个容量为 n 的样本，并用一个量度指标对它进行测量（如问每一被调查人一个像“过了最近一个生日之后，您的年龄多大了？”这样的问题），我们便可以得到一个总体参数的估计值，在这一例子中，它就是平均年龄[3]。如果调查的其他方面的条件保持不变，我们又选取了第二个容量为 n 的独立样本，采用同样的量度指标，我们便会得到平均年龄的第二个估计值。如果我们将这一程序继续进行下去，我们会看到不同的样本之间存在的平均年龄的变差。我们同样也可以重复调查的其他程序，当然仍需保持其他的程序不变，我们同样可以预期，在诸如这样或那样的其他程序的重复过程中，也会出现类似的随机变差，如调查可以联系到的被调查人的百分比，或拒绝回答有关个人收入方面

的问题的被调查人的人数等。任何一次特定试验的变差量,既可能高于,也可能低于各次试验的变差量的平均数。

类似的效应也可以来自调查问题的本身。设想我们对一个被调查人的样本提出一个“您每月购买日杂用品多少次?”这样的问题。无疑将会得到一个样本估计值,譬如每月 3.4 次。现在,假定在一星期之后,我们再对同一组被调查人提出同一个问题。许多被调查人给出的答案可能与上次相同,但某些被调查人的答案则可能有所不同,稍微高一些或稍微低一些都有可能。导致这样的现象的因素可能是多方面的,如只是因为在第二次调查时,被调查人在回答时考虑得更加仔细或更随意了等。但是关键在于它与不同的样本可能产生不同的估计值一样,对同一样本用同一问题进行再次调查,因为调查实施上的差异,估计值同样也会有所差异。在这两个例子中,如果这些差异是随机的,那么我们就把它们看作一个方差源。

与方差不同,偏倚发生在量度倾向于比总体值恒高或恒低时。在我们刚刚引用的那一例子中,对某些被调查人来讲,他们给出的答案可能倾向于比实际年龄高。例如这有可能是问题措辞含混不清的结果。假定我们问题的措辞是“您的年纪多大了?”那么大多数被调查人将会给出他们目前的年龄——就是他们在过了最近的一个生日之后的年龄。但是另一些人,可能认为他们的生日已经快到了,他们应该报告他们即将要到的年龄。一般讲,没有一个被调查人会报告自己以前的年龄,即使他们刚刚过了生日。正因为这样,只要错误报告的发生是非随机的,那么某一错误发生的方向就不可能与另一错误相反。这样的错误报告的年龄可能高于实际的年龄。因而由此得到的样本调查的平均年龄的估计值将高于真实的均值。在这一例子中,我们便说这一量度是偏高的[4]。注意,这是除方差之外的另一种误差源。而更多的情形是,特别是在小规模的调查研究中,除了一些重大的例外之外,我们不(或在我们资源许可的范围内,不能)在一个特定的调查中,对各种误差源使用经验性的量度。我们的设计决策与非抽样误差有关。这些决策的主要参考资料来自其他的调查发现和曾经使用诸如这样的量度做过的试验。我们使用从以前的方法学研究中得到的、更为普遍的(但希望是比较稳定的)发现,作为寻找在我们自己的调查中可能存在的误差源的指导。

在一个将调查研究的状况作为一门科学的讨论中,格罗夫斯(Groves,1987)把调查的误差问题归结为人们在测量和减少两个方面所做的努力。这就是说,那些“试图构建调查误差的经验估计量的人和那些试图消除调查误差的人”做出的种种努力。合乎逻辑的观点是,研究者应该在量度和减少二者之间持平。不仅如此,为了减少主要的误差源,研究者所做的努力必须有一定的导向,并以经验的估计为依据。但在很多场合,情况却并非如此(由于某些原因,我们无法在这里展开这方面的讨论)。而以下的那些讨论主要集中

在如何减少数据收集过程中的误差，如何确定误差源，及如何减少误差在调查结果中的效应等问题上。

调查质量的量度

数据收集期间产生的误差可能导致潜在的严重问题。例如，如果没有回答问题的被调查人与回答了问题的被调查人存在差别，那么由调查得到的估计值就会有一定程度的偏倚。当然，在一般情况下我们并不知道二者在调查的量度上是否存在着差异，因为由定义可知，我们没有来自没有回答问题的被调查人的数据。在缺乏调查数据的时候，我们常常参阅一些间接的指标。例如，在一般人口的调查中，我们常常从普查的数据了解数据的某些人口学特征的预期分布，如年龄、性别、种族和教育等等。如果像正常的情况那样，在我们的调查中也收集了某些诸如这样的信息，那么我们就可以将调查到的被调查人的人口学特征与普查的数据做一个比较。假定我们在比较之后发现：①我们的被调查人的某些人口学特征的分组人数过少；②来自人数过少的那些组的成员与非那些组的成员对很多调查的问题答案不尽相同。那么以那些人数过少的组为参照，我们有理由担心我们做出的估计可能存在偏倚。如果一个组别的代表人数比例过低，那么某些其他的代表人数的比例就必定过高，懂得这一点是很重要的。我们不妨来考虑一下，情况为什么会这样？

数据收集工作完成之后，但如果尚未开始分析工作，我们可能不知道人口学特征和许多实质性变量之间的关系（当然如果它们之间存在关系的话）。避免这些潜在的偏倚的唯一可能是争取达到尽可能高的回答率。所谓回答率是合格的已经做完调查的样本单位的百分比。它也叫做单位（unit）回答率。后面我们将会对这一问题进行更为详细的讨论，其中包括它是如何计算的问题。一个与之类似的量度指标叫做合作率（cooperation rate），它是接受调查的样本成员与接受调查的人数和拒绝调查的人数之和的百分比。我们来考虑一下这两个比率有什么不同。

当然，理想的情况是，我们希望能从所有选取的被调查人那里，得到所有的问题的答案。我们知道我们可能在两个层次上偏离我们的目标。一是在单位的层次，即在个人或住户的层次（但是它也可能是一个组织机构、商行或学校，如果它们就是我们调查的总体的话）；二是在题项的层次，即问卷中的一个单独的问题。数据的丢失无论发生在单位层次还是题项层次，都有可能对我们调查的数据的质量造成问题。如果我们无法从某些被调查人那里得到任何信息，从另外一些被调查人那里得到完整的信息，那么我们的估计和其他各种分析就可能不够正确，有时甚至会很不正确。

单位回答率是主要的，也是最为人们所普遍接受的质量的量度指标。不言而喻，同意接受我们调查的被调查人可能并不一定会回答所有的问题。他们可能拒绝回答某些问题，或在不经意间漏答了某些题项，而这些都可能引起题项的无回答。尽管我们认为，我们对于题项无回答问题的关注应该无异于单位无回答问题，但是人们通常对于这一问题的关心似乎很不够。在一般情况下，那些同意接受我们调查的被调查人将会回答我们的全部问题，或几乎会回答我们的全部问题。然而，如果调查的问题比较敏感（如性行为或非法行为），或令许多被调查人都觉得难以回答，那么题项无回答问题就会变得很严重。题项无回答问题一般都集中在少数几个问题上。

两种类型的回答问题都可能受到调查员调查方式的影响。调查员的行为可能会对被调查人参加调查的态度和他们回答某些特定问题的态度发生影响。调查员也可能对被调查人给出的答案的信度发生影响。这就是说，调查员自身有可能成为调查结果的可变性（variability）的一个来源，例如，对开放式问题答案做了错误的编码或在传达处理问题时前后不一致等。正因为如此，调查员的培训和督导对高效的数据收集是十分关键的。调查实施情况是难以在数量上加以测量的，即难以把它从加总的回答率中单独分解出来。但是，调查的执行情况的确是调查方差和偏倚的一个来源，这一点是毫无疑义的。

单位无回答

我们之所以要关注单位无回答问题是因为它常常是由信息模式的丢失引起的。例如，假定在犯罪问题调查中，在单位层次，样本成员中的男性，或文化程度较低的成员，或居住在郊区的老人比较可能采取不合作的态度。如果这些样本成员和与合作调查的成员的态度或经验有所不同，那么我们旨在以调查结果代表州的整个成年人总体的目标将会因此而受到影响。例如男性较之女性更不愿意考虑判决的可能选择；文化程度较低的人对警察工作的评价可能低于其他人；而老人则更可能因为担心遇到罪犯而避免去某些场所等。所有这些情况都是我们推测可能发生的情况，但是诸如这样的模式都有可能在调查结果中发现[5]。各个子群体之间在观点和行为上存在的差异、子群体的代表人数超出比例或低于比例，都会在一定程度上对调查的结果有所影响[6]。

在 20 世纪 70 年代，电话调查变成了收集一般人口总体的数据的主要方法（在联邦政府之外），有鉴于此，迪尔曼（Dillman, 2000）指出，必须对整个调查过程中的每一个步骤详加检查，因为它们都关系到调查最终是否能够成功，其中回答率的高低，对调查最终是否成功的关系尤其重要。今天我们再一次回顾这一忠告是十分有用的。它告诉我们，在进行抽样调查的时候，我们既要考虑到新的调查工具的长处，也要考虑到它们的不足和障碍。

如果我们有条理地考虑一下调查的实施过程中所有可能对被调查人参与调查的决策产生影响的要素，我们就会认识到某些很重要的要素，可能在我们制定调查计划（表9.1）的时候，已经在不经意间被我们忽视了，或虽然注意到了，但却注意不够。例如，我们注意到某些要素在调查员与被调查人接触之前，已经到了被调查人那里，而它们提供的信息对被调查人是否参与调查的决定有所影响。

有时，在一个电话调查开始之前，我们会给那些把电话号码刊登在电话簿上的样本户先寄去一封信件。作为随机被调查人选择步骤的一个组成部分，调查员常常会在与被调查人接触之前，与户中的另外一个成员交谈。而这个成员对与我们交谈的印象都会对被调查人最终是否能与我们接触有所影响。即使被调查人本来就愿意参与我们的调查，或者甚至最初就对我们的调查有一定了解，这一影响也是不可避免的。有时我们虽然没有和户中任何一个人交谈，只是在他们家里的电话录音上留下有关调查的一些信息，在信件或留言中留下了可以获得有关调查的信息的联系电话或网址。即使是这样，在很多时候，人们仍然可能对我们的电话不予理会，尽管我们已经让其知道了我们的联系电话或在来电显示系统中留下的组织名称。

这些先于调查迈出的每一步都可能会影响被调查人的合作和/或数据的质量。那么对于这样的问题我们可以采取一些什么措施呢？如果我们把自己变做一个可能的被调查人，从这一角度来考虑一下这些问题，可能对问题的解决不无帮助。我想这是一个很有用的练习。在正式开始电话调查、邮寄式调查或互联网调查之前，先来编写一个详细的数据收集的方法步骤（下面的表9.1将会对方法步骤的编写有所帮助）。编写好了之后，与一些同事就这些步

表9.1　影响单位回答的因素

事先通知将要进行的调查
努力尝试与被调查人联系
初次联系和经（看门人）中介的被调查人选择
请求被调查人参与调查
继续努力联系
拒访转化

骤中那些将对被调查人参与调查的态度有所影响的问题进行讨论，听取他们的意见。做这一练习的一种方法是，让你的同事“扮演”被调查人，而你们自己则“扮演”研究人员。在调查的每一阶段，我们都要制定详细的实施计划，例如预先发出的信件的内容包括什么；我们是否要在应答机上留信息；如果留信息，留些什么；在与户中第一个人交谈时，如何介绍我们的调查等等。在对调查的每一个阶段的操作步骤做了介绍之后，我们的同事——被调查人就会

告诉我们,他们的反应是什么。这些计划好的方法步骤的效应是正面的还是负面的?在我们的同事扮演的被调查人心中出现的问题或疑问是什么?每当我们得到的反应是负面的时候,我们都应该考虑对原来的方法步骤做一定的修正,然后再在“被调查人”身上进行试验。我们将会发现,不需要任何专门的知识,只要你和你的同事能周密地考虑这个过程,即使是以常识的方式考虑,都会对如何最佳地进行数据收集这一问题——从被调查人的角度——产生许多思想和洞见,并对这一过程的关键有所了解。

有若干种可资利用的、用于防止单位无回答的方法可供我们选择。它们既可用于调查员填写的调查,也可用于被调查人自行填写的调查。这些方法包括设计令人感兴趣的、结构合乎逻辑的和简单明了的问卷;采用有效的方法提供有关调查的信息,如事先发一封信件(电子邮件)和精心编写的简介或封面信件。其他的那些关键的设计因素只适用于这种或那种调查方式,而非所有的调查方式。

对于使用调查员进行的调查,认真地培训调查员是很关键的。在被调查人自行填写的调查中,被调查人使用的填写说明必须绝对清楚和易于操作。

近年来无回答呈上升趋势

自从 20 世纪 90 年代以来,若干社会和技术的因素曾经对数据收集,特别是对一般人口总体的数据收集有影响。有时候,新技术(如互联网和其他计算机辅助的数据收集法)曾经为低成本的收集创造了潜在的机会。而在另一时候,诸如拒接来电这样的技术给数据收集造成了很大的困难。只用作互联网的线路、传真机的电话号码比例的上升,以及至今还仍居次要地位的手机使用率的上升,使得抽取通向住户家中的电话号码更困难和费用更昂贵。

电话推销的迅猛增长,使许多可能的被调查人觉得难以将它们与合法的调查电话加以区别,因而也成为了一个妨碍调查进行的严重问题。许多住户为了防止轰炸式的电话推销所采取的措施对合法的调查有着决定性的影响。

电话调查回答率的下降便是这些问题可以测量的明证。与十年前相比,甚至与四五年前相比,被调查人已经变得更加难以联系和更不愿意参与我们的调查。在回顾联邦政府资助组织的几个调查的无回答问题的发展趋势时,格罗夫斯等人(Groves et al. ,2004)发现无回答问题呈现一种一致的、虽然不是剧烈的上升的趋势。人们普遍认为,非联邦政府资助的调查情况也并不乐观。

至今我们仍然不知道回答率的下降,究竟在多大程度上影响了调查的质量和调查者对自己调查结果的信心。但是毫无疑问,回答率的持续下降最终将可能是十分有害的。

调查研究者对这一趋势采取了许多应对的措施。虽然我们无须将那些保

持回答率的新方法和那些一直沿用至今的传统方法截然分开，但是了解调查的这些方面的工作在调查资源中所占的份额已经普遍增长这一点却是很重要的。

题项无回答

数据可能会随机地丢失。这就是说，任何一个问题都会像其他任何一个问题一样有可能被丢失。随机地丢失数据可能是某些错误，如调查员的错误、被调查人的错误（如在下面所提到的填写邮寄问卷时进行跳答题），甚至编码错误的结果。如果随机丢失的数据数量不是太大，那么它们对我们的调查结果的影响不会太大。当然，如果丢失的数据只有几个百分点，我们大可不必为它的影响而担忧。在这样的情况下，我们不必动用很多资源来降低或消除随机丢失数据的问题。遗憾的是，在数据丢失时，大多数丢失都是非随机的。

即使那些由于被调查人或调查员的错误而导致丢失的题项也可能存在某种模式。在自行填写的调查中，被调查人最有可能掠过或不回答那些含糊不清的、敏感的或难以回答的问题，或在前面有含义不清的引导语的指示下跳答题。调查员可能在不经意间对那些不合自己意愿的问题，鼓励被调查人不回答（或对答案施加影响）。我们需要在问卷设计、测试和调查员的培训期间，对诸如这样的潜在的问题加以必要的注意。

我们的措施：决策和方法

下面我们将讲解我们可以采取的措施。这些措施都是针对数据收集过程中的特定误差源的。我们的措施集中在两种类型的决策。一种决策类型与设计有关，另一种则与设计的实施程序有关。就决策而言，我们设计的问题主要包括数据收集方法的选择、起草调查员培训方案和制定数据收集的计划等（包括回拨和拒访转化）。这些设计决策不仅是不可或缺的，而且在一旦决定之后，是相对固定的。在调查过程中，我们无法轻易地从邮寄式调查转换到电话调查，或专门增加一轮拒访转化的工作。这些决策都必须在最开始的时候做出，它们决定了数据收集的主要特点。而数据收集阶段的费用几乎总是调查中最高的。因而对它们进行任何的改动，都必定要付出高昂的代价。

在这里我们将程序看作调查实施的一个要素，它虽然它建立在调查之初，但却是可以在调查过程中加以改动或调整的。这些程序要求在调查过程中进行动态的监控和微观的决策，如对有问题的调查员或被调查人的行为做出处理，并解决在争取合作过程中发生的问题等。

在某种意义上讲，设计和程序之间的区别是人为的。不仅如此，我们认为

这种区别将为我们提供一种有用的框架，用以区别那些必须预先完全确定的主要设计决策和那些在调查实施过程中吸引了我们大部分的注意力的、随时进行的微观决策。

数据收集程序的选择除了必须适合我们有关调查做出的决策之外，它还必须受到我们预算的限制。每一种选择都需要一定的代价，在这里代价的含义是广义的，它既包括金钱，也包括其他的资源，如研究者可资利用的时间、无酬（自愿）的助理或同学。即使我们有某些可资利用的“免费的”资源和设施，但它们一般也都是有限的，因而也需要就如何使用这些资源做出决策。例如，虽然一个大学老师做的调查可能不需要为秘书打印的培训材料付费，但是秘书的工作时间并不能全部被我们的项目占用。这样我们就必须在工作安排时有所选择——究竟安排秘书打印培训材料、处理预发信件还是做其他的工作，以最佳地使用可资利用的时间。最后，为了解决某些特殊的问题，我们必须在调查期间重新配置我们的资源。

无论调查是否是个别进行的，也不论调查是通过电话、邮件、互联网或其他方法进行的，数据收集过程都需要按规定的程序进行，如发邮件、拨样本电话号码、建立样本文件和跟踪样本结果等。这些都是相当简单的程序，它们都有可能产生误差。我们必须保证拨叫的电话号码正确无误，并保证通过样本跟踪使所有样本都能正确地发放用于数据收集。

我们可以通过例行程序的建立，保证万无一失地完成这些工作量相当大的书记任务，并可防止将严重的误差引进我们的调查。调查的这些部分的工作大部分都是保持记录这一类的工作。许多计算机化的数据收集的工具，如CATI（计算机辅助的电话调查）的许多功效都对这一类工作的完成很有帮助。我们将从调查必须执行的每一个步骤来考虑这一问题，并为每一步都设计了记录保持的方式。这些看似平淡无奇的程序，将确保我们自己不致未给某些未回答问题的被调查人寄发补充调查的邮件（或给那些已经完成了调查的样本成员，或已经将问卷寄回给我们的样本成员寄补充调查的邮件，造成资源的浪费。）、未给每一个电话号码以相同的试拨机会，或未能再次对那些并没有坚决拒绝接受调查的被调查人做工作[7]。所有诸如这样的问题都能导致误差的产生和资源的浪费。然而，只要我们对这些问题的详细资料稍加注意，就可以将这些误差源减少到很小的程度。在数据收集期间，我们可以诉求的主要资源是我们上面介绍的那些可以减少严重的调查误差发生的活动和行为：单位无回答、题项无回答和调查员及被调查人效应。

由调查员填写问卷的调查

被调查人参与调查的决策可能会受到很多因素的影响[有关某些为这些因素建模的阐述]请参见格罗夫斯等人的著作(Groves et al.,2004)。我们只能对这些因素略加控制,或完全无法控制,例如调查的主题、调查的资助者、在任何一个调查中影响被调查人参与我们调查的主观原因或调查进行的社会环境等。我们力求调查的主题能为人们所关注和感兴趣,并再三强调调查的重要性,但是在不致误导我们的调查的前提下,我们在这些方面的作为可能只有那么多。

我们可以控制的影响调查答案的因素是调查的实施程序,在某种程度上也包括调查员的行为。我们力图设计一些程序,以使调查的每一步(见表9.6)都能以最佳的效应被执行。然而,那些程序,如上佳的预达信件、撰写得当的调查简介和适当的物质奖励,都可能在与被调查人交流的过程中,被那些缺乏训练的调查员所破坏。一个实施得当的调查必须兼顾数据收集程序和调查员效应二者。

我们首先介绍数据收集的程序,然后再讨论调查员的问题,讨论主要围绕电话调查进行。尽管高回答率主要都出现在面对面的调查中,但因为费用的原因,这样的调查普遍程度远远不如电话调查。我们的讨论也涉及某些面对面的调查,因为存在着某些例外,譬如特殊人口总体的调查。在个别调查数据收集中,有时也需要采用多种方式的调查。

重要的问题是我们必须将我们的资源用于最有效益的地方。在无论哪一种类型的调查员填写的调查中,所有致力于降低单位无回答率的工作都是体力密集型的,因此费用是比较高的。这一情况对电话调查和个别调查都是如此,尽管后者的费用大大高于前者。在无论哪一种类型的调查中,有关扩大合作率的程序的决策将对我们有限的资源有很大的影响。我们的讨论主要围绕电话调查。讨论的问题一般适用于任何类型的调查员填写的调查。对那些只适用于个别调查的问题,我们则以适当的方式另行评论。

调查实施的程序和调查过程中的质量控制一般都通过计算机辅助的数据收集和数据管理工具实现的。这种方法的英文缩写是CASIC[计算机辅助的调查信息收集(computer-assisted survey information collection)]。它涉及一系列广泛的、用于收集调查数据的计算机驱动的技术[8]。两种主要的CASIC工具是CATI和计算机辅助的个人访谈(computer-assisted personal interviewing,CAPI)。这两方法都自20世纪70年代开始使用,一直沿用至今,它们都允许调查员在进行调查的同时,将调查得到的答案输入计算机文件。这些系统的

主要优点是进行质量控制。CATI 和 CAPI 系统可自动处理问卷的跳跃模式；降低调查误差；对某些题项允许输入的数值范围加以限制，例如在成人调查中，要求年龄这一条目是一个两位数；通过与其他题项的答案比较来检查某些答案的内在一致性，例如提醒调查员，如果某些出生于 1970 年的人报告说，他们于 1980 年高中毕业。这些技术的可贵之处在于它们能在调查的同时，而不是在调查之后的数据清洁阶段实现这些功能，从而能有效地防止为清理或补充数据与被调查人进行费用不菲的再次联系。

许多这样的系统也可自动地执行某些样本管理和调查进度安排的任务。大多数专业的和大学的调查组织都有某种类型的 CATI 系统，而 CAPI 系统与个别调查一样，使用不太广泛。

如果在我们的调查中有可资利用的 CASIC 系统，那么就应该将学习如何使用这些系统，包括将问卷编程输入系统所需要的时间和费用，都加入到我们的预算和项目计划中去。对于一次性的调查，除非使用的系统非常简单，最好还是另聘专门的人员进行编程，而不要去尝试自行编程。

这些工具对调查过程的跟踪、数据的检查和在数据收集过程中及时修正某些程序是非常有用的。例如，在计算机辅助的条件下，如果我们认为某种变动将会使筛选变得更精确或更有利于被调查人的合作，我们便可以马上更改住户筛选程序。同样，如果我们认为调整回呼计划的确更有效率的话，我们也可以及时调整回呼计划。

我们可以检查数据的无回答模式，例如我们是否在某些地区（如大都市地区的中心城市）或某些子群体（如青年男性）感到特别困难？将我们的数据与普查数据（面对面访谈或只是电话住户调查的总体）的分布做比较，我们可以清楚地发现诸如这样的不成比例的答案。如果我们发现有诸如这样的单位丢失的模式的确存在，我们也许就应该把我们的注意力转向那些代表人数过低的群体。采取的方法既可以是给这些群体分配更多的总调查时间，也可以是安排特别有经验的调查员与这些群体的样本成员进行联系，或者二者同时并举。如果我们发现在某些地段进行调查特别困难，那么我们就可以把这一地区个案更多地分配给更有经验的调查员。如果我们发现男性被调查人的拒访率比较高，那么就可以把这样的个案分配给善于做拒访转化的调查员。另一种可能出现的模式是不成比例的单位无回答都是某些调查员所致。如果情况的确如此，我们必须重新培训这些调查员，或重新分配这些调查员的工作[9]。

在讨论数据收集的实施步骤时，重要的问题是我们必须牢记 CASIC 技术的能力和可行性。

事先通知

在开始拨电话号码(或当面调查)之前,我们可以先发一封信件,通知样本户他们已经被选作了我们的调查对象。并对调查的内容、组织实施的机构和调查目的做一个简单介绍。诸如这样的信件应该对保密性做出承诺,并提供一个联系的电话号码(或者互联网上的网址),以便被调查人与我们联系,进一步了解有关调查的各种信息。信件当然必须简短,但除此之外,它还必须提供有关调查的基本信息,给可能的被调查人解释调查为什么如此重要。

现在我们再回到马里兰犯罪问题调查,考虑一下我们应该怎样就这一调查构建一封预达的信件? 在这里我们之所以使用“构建”一词,是有目的的。这样一封信件应该由几个方面的内容组成,每一方面都阐明一个可能对可能的被调查人是否参与我们的调查的决策有所影响的若干种因素中的一种。如果被调查人根本没有看我们发给他的信件,尽管我们用心良苦,结果也是于事无补。为了使被调查人能看我们的信件,我们必须尽力使预达信件不仅简短,而且易懂。信件的内容仅限于那些对被调查人的态度最有影响的问题,且这些问题还必须表述的简单明了。

这一信件由几个简短的段落组成(表9.2)。某些段落只是简单扼要地介绍一下调查项目,另一些段落则说明一下调查的重要性,还有一些段落可能需要对调查的组织者和调查结果的用途做一些介绍。当然,最后不要忘记写上与我们进一步联系的方法。作为一个练习,我们来考虑一下这一信件中的每一个句子,每次只考虑一句。应该考虑的问题是:写这一句子的目的是什么?这一句子在争取得到被调查人的合作过程中,重要性如何? 我们再一次提醒大家,在考虑这些问题时,都应力求从被调查人的角度,设身处地进行考虑。如果我们自己收到了这样一封信件,我们的反应会是什么? 它是否会促使我参与即将进行的调查?

不言而喻,预达信件都需要一定的费用。但是否发送预达信件的决定权却在我们。不过无论它是否值得,它所需要的费用和时间都难以完全确定。不仅如此,信件也只能发送到那些刊登了电话号码的住户。(出售样本的卖主也可以提供住址。如果样本是我们自己选取的,我们可以把它寄给一个经营商业数据库的卖主,要求他为我们提供样本的住址。)

只要条件允许,我认为还是事先发一封信件为好。这样的信件将增加调查的合法性,并使它区别于电话推销,而这些对于我们的调查来讲都不是小问题。事先发出的信件还可以有其他的作用。对那些未曾刊登电话号码,并对调查的合法性有疑问的住户,如果这些可能的被调查人能够提供他们的住址,

调查员也可以给他们发这样一封信件。如果我们事先已经准备好了这样的信件,这一过程的实现便会快很多。

表 9.2 预达信件

Jones family
2189 Cedar Rd
Adelphi MD 20799(系姓名住址,保留原文,未译。——译者注)

亲爱的住户——琼斯和您的家人:

在最近几天您可能会接到我们的电话,希望您家能参加一个重要的调查。马里兰大学正在为州最高当局做一个有关暴力犯罪的调查。

您的家庭已被随机地选入这次调查。本次调查希望能代表马里兰所有类型的家庭,包括那些已受犯罪之害和尚未受犯罪之害的家庭。您给出的所有答案将会严格保密,且调查的结果只会报告给有关组织团体。

与暴力犯罪有关的最高当局将用这些调查结果帮助制定全州防止犯罪的规划。

您是否参与我们的调查是自愿的。但它对我们的调查的成功是很重要的。调查可能需要 10 ~ 15 分钟。如果调查员刚好在您不方便的时候给您打电话,她会很高兴在您认为最合适的时候,再一次给您打电话。如果您对我们的调查还有任何问题,您可以拨打 1-800-314-9090,或登录我们的网站www. stopcrime. umd. edu 查询。为您将要为这一重要的项目所花费的时间,预达我们的谢忱。

谢谢!
马里兰大学
调查研究设计系
罗伯特·爱丽思博士

联系样本被调查人

如果许多选出的被调查人都没有联系和调查到,那么我们在设计和抽取概率样本时所做的种种努力和花费的精力都将付之东流。我们已经注意到丢失数据可能产生的问题,并对调查员和被调查人的行为对这一问题也会有所影响这一点有所了解。现在我们转向回拨程序的使用问题。很多调查告诉我们,不论什么调查方法,反复回拨对提高选取样本的回答率都会有很大的影响。回拨是一个降低无回答的最有效的策略。

无论抽样框是随机数码拨号(RDD)、一张清单,还是有关信息的组合,在第一轮拨叫之后,样本都会被分成若干个不同类别的组。这些组包括那些已经完成调查和已经可以确定是非住宅而无需进一步注意的住户。其中也包括较早的拒访户、那些(住宅或非住宅的)号码的具体状况还不能确定(铃声响,但无人接听)的号码和一些被调查人(或户中的接电话的人),因为诸如听力

或语言这类问题而无法接受调查（至少是在第一次联系时是这样）。此外，我们必定还会有数目众多的未曾联系者，其中包括接通了应答机或发现随机选取的被调查人目前不在家或没有空等。

对于样本号码的每一种初始处理方式，我们都必须决定随后进行的回拨的策略。这一工作的好坏对我们是否能与选取的被调查人取得联系，并进行调查有很大影响。当然，它也对采取这一措施所需费用的高低有很大影响。

在电话调查和当面进行的调查中，我们会有某些有关样本单位在第一次拨叫时为什么没有调查成功的信息，当然，这些信息充其量无非是一些我们在试拨时得到的那些信息，即使如此，我们也可以把它们用来拟定我们的回拨策略。但是必须指出的是，只有在调查员试拨时仔细地记录了当时的情况，我们才可能有这些信息。

调查员必须进行培训，学会记录表格和程序的使用，以对每一次试拨的结果，即一个样本号码的每一次拨叫的结果做记录，并在样本状态报告中对记录的结果进行加总。不论拨叫的结果怎样，调查员都必须如实对拨叫时实际发生的情况编码。诸如这样的编码可以相当详细，但是我们认为并非越详细越好，只要能满足研究的需要就可以了。最基本的拨叫结果包括完成调查（completed interview）、最初拒访/完成部分调查（initial refusal/partial interview）、最终拒访（final refusal）、无法联系（noncontact）和其他无回答问题。调查员应该对这些处理结果的每一个的具体日期、钟点和星期几，一一记录。

样本中每一号码的初次拨叫的结果将决定我们随后应该怎样来“操作这一样本”。这就是说，我们必须首先确定是否要继续回拨，如果要，那么另外的拨叫应该怎样计划安排，而在有的时候，还必须考虑由哪些调查员来拨叫。我们这些策略的好坏，将对调查费用高低和非抽样误差的大小都会有很大影响。因为对样本号码的每一次拨叫都需要一定的费用，我们的目的是用最小的代价达到我们对每一个电话号码做出部署安排、确定合格的样本号码和完成尽可能多的调查的目的。这些策略的好坏之所以对费用有影响是因为，总预算中相当大的一部分是用于数据收集的，而拨叫样本号码的费用即电话费，则在实际访谈费用之后（译者注：调查员报酬），在数据收集的总费用中占第二位。

表9.3是初次拨叫结果的分布（对每一电话号码拨叫一次之后）和马里兰大学调查研究中心1992年州犯罪问题调查的实际最终分布。第一列相当于我们的第一个样本状况报告。在拨叫一次之后，相当数量的非住宅号码便得以确认，但是更多的号码则需要随后再次拨叫。最初我们无法确定一大堆电话号码究竟是不是住宅电话。但在随后再次进行拨叫之后，这一类的数目大大降低了（尽管还没有减到零）。在确认的住户中，有60%以上（835户中的527户）选定的被调查人在第一次呼叫时未能调查成。然而，在调查结束的

时候,无法联系的比例已经降到了 5%。

表 9.3 马里兰犯罪问题调查:一次试呼和 20 次试呼之后的样本处理结果

	试拨一次之后		试拨 20 次之后	
总样本	1 816		1 816	
非住户	545		702	
住户状况未知	436		53	
住户	835	100%	1 061	100%
访谈	181	22%	824	78%
拒访	54	6%	164	15%
无法联系	527	63%	49	5%
问题户(语言、听力、疾病等问题)	73	9%	24	2%

让我们来对每一类处理结果逐一进行考察,看看在第一次(和随后的)拨叫之后,有什么样的信息可以收集,进而看看,我们可以怎样来使用它们。有些样本号码几乎马上就可以删除,因为它们并不属于合格的住户总体:营业性号码、未运行号码、无法接通的号码、政府办公室号码等等。不过,即使这样一些电话号码,有些也还需要再次进行拨叫。

第二组人是那些很快就可以了结的人。这些人很容易联系,且马上同意接受调查[10]。一般讲,这些人是那些经常在家的和同意马上接受调查的被调查人,他们大多是老年人、退休的人、失业的人,或在数量上低于前几种人的妇女。虽然这些人都是合格的样本成员,但是他们并非总体的随机样本。使用这些样本,我们将可能得到偏倚非常严重的总体估计值。存在若干种非常有诱惑力,但却是不正确的程序,这些程序最终将导致诸如这样低质量的样本。例如,我们可以抽取一个相当大的电话号码样本,立即开始拨电话进行调查,而在预定的调查数已经完成的时候,马上停止调查,终止拨叫剩下的其余号码;或者,我们可能对每一选取的样本号码试拨一次或二次。这两种方法都会只调查到那些最容易联系和进行调查的被调查人,从而导致具有上面列举的那些人口学特征的群体人员在样本中的比例过高。

另一种导致某些被调查人马上同意接受我们的调查的因素是调查的题目。那些对调查的题目感兴趣的被调查人很少会把调查员打发走。在马里兰犯罪问题的调查中,这样的被调查人可能是那些身受犯罪之害的人,或因为其他的原因(如因为他们的职业或朋友的经验)对犯罪问题有强烈感受的人。同样,他们虽然都是合格的样本成员,但是一个只包含了这一类型的被调查人的调查,其提供的结果可能是错误的。两种类型的被调查人都很容易联系,也

都不需要进一步的要求或做工作就会很快同意接受调查，但是我们也同样可以证明，这两种类型情况都使我们的样本具有便于调查的子群体性质，而不是随机的。现在问题已经变得很清楚，从样本户到调查完成的转化并非一个随机过程，我们必须引导我们自己的工作（资源），密切注视已经完成的样本的构成比例，并最有效地将那些不合格的东西清除出去。

在第一轮拨叫之后，剩下的样本成员被分为两种类型：住户和住户状况未知。在我们处理这一样本时，同样的模式将会反复出现：某些被调查人比较容易联系和调查到，另一些人则需要做很多工作才能联系到，或要费很多口舌才能说服他们接受我们的调查，或者集两者于一身，既难以联系也难以说服接受调查。

正如我们在第 7 章已经讨论过的那样，一般情况下，我们在开始数据收集的时候，先发放一个电话号码样本的随机子样本，一一拨叫。在根据这一子样本估计了完成率（已经完成的调查数除以最后确定的样本数）之后，我们再发放另一些随机子样本，以使样本量最终能达到我们的调查预定的总调查数。在把这些号码发放给调查员的时候，呼叫结果的模式一般大同小异。

联系次数（number of contacts）。能否成功地与选取的样本成员的大多数取得联系主要取决于试图与他们的联系次数。那些只尝试一次联系的调查，几乎都可能会有严重的无回答偏倚。就电话调查而言，尝试的次数可能在 3 ~20 次，而邮寄式调查则在 2 ~4 次（Groves，1989）。我们的认为，一般人口总体的电话调查应该不少于 5 次，而邮寄式调查则不应少于 2 次。如果在数据收集期间，我们发现预定的努力程度不能得到预期的回答率，那么我们可能需要对我们原定的回呼标准做一些修改。遗憾的是因为费用的原因，我们并不具有做过多的额外联系的自由。但是对所有没有联系到的样本成员增加一或二次额外的呼叫，一般都会是很有用的。如果在考察了每一个成员过去的呼叫结果之后，能对那些号码在尚未对它们呼叫过的日子和时间，再次集中进行呼叫，可能是会很有效果的。

应答机（answering machines）。近年来，越来越多的家庭都有了家用应答机。早期进行的有关研究（Oldendick & Link，1994；Triplett，1994）表明，大多数这样的住户不必事先做什么计划都可以联系到。周末的早上与这样的住户进行联系特别有效，且在一旦取得联系之后，与那些家中没有应答机的住户相比，他们似乎显得更加合作。在犯罪问题调查中，第一轮拨叫之后，在确定为住户的个案中，大约只有 13% 左右作为应答机处理。在调查结束的时候，这一数字降到了 2% 。尽管这一呼叫策略至今仍为我们所推崇，但是有某些证据表明，它的效率已有所下降。在一项目前正由疾病控制中心主持进行的大规模的、每季度都选取独立样本的免疫调查中，2002 年，在数据收集结束时，处理码为应答机的百分比开始上升。

约会(appointments)。在电话调查中,我们常常会遇到这样的情况,那就是被调查人要求在他们比较方便的时间再与他们联系,并愿意为调查定一个约会[11]。有条理的组织调查工作,以保证安排好这样的约会,做到这一点是非常重要的。错过了这样的约会,一个本来愿意与我们合作的被调查人很容易就会变成一个不合作者。

拨叫时间安排(call scheduling)。为了恰当地和以较高成本效益处理样本,调查员必须精确而完整地记录每一次试拨的结果。而这一目标是借助填写一张简单的,但必须在拨叫每一个号码之后立即填写的表格实现的。这一记录使调查人员的管理者得以在考虑对每一个住户进行额外的试呼时寻求恰当的模式,确定什么时间做再次试呼合适,什么时候不合适。例如,假定某一特定的住户,在平日黄昏的时候,从来都没有接过电话,那么我们就应当将后几次试呼的时间改为周末的晚间。户中的知情者,一旦联系上了,他们很可能会为我们提供与随机选定的被调查人联系的最佳时间。这一信息应该记录在同一张表格上。

用一个具有日程安排性能的 CATI 系统来安排日程,比用人工安排日程更有效。不过这需要事先做出决定,并把这些决定输入 CATI 系统。我们需要决定第一次拨叫应该在什么时候进行——白天还是晚上、平日还是周末,以及在拨叫的结果是未能实现调查时,根据呼叫的结果是有铃声没有人接、拒访、应答机或其他情况,对随后进行的回呼做出相应的安排。

一般讲,就 RDD 调查而言,将第一次拨叫安排在白天是比较高效的,因为这时,系统可以把营业性或其他非住户的号码去掉。此后,平日的晚上和周末则比较好,因为这时大多数人可能都在家。把回呼安排在不同的日子这一点很重要。

确定坏号(identifying bad numbers)。在确定某些电话号码是否是住宅电话号码时,我们可能会有一定的困难,即使试拨多次,也未必能确定[12]。因为电话系统的技术方面的缘故,一个电话被拨通之后铃声连续不断地响起来的号码,可能是一个住宅号码,也可能不是住宅号码;可能是一个正在服务的号码,也可能未在服务中的电话号码。反复地拨叫一个非住宅号码,不仅浪费时间和金钱,而且在采用梅托富斯基—沃克斯博格样本设计(二阶 RDD)时,将会使我们丧失替换号码的机会。为了解决这一问题,有三种方法可供我们选择:第一,在比通常情况早一些或晚一些的时候,计划安排打几个电话,试图能联系到那些工作日程异于常人的人。如果这些号码被证明的确是住宅号码,并进而与被调查人定下了调查的约会,那么安排在这些零星的时间拨打的电话将会对降低偏倚很有用处。诸如这样的被调查人可能在人口学和其他实质性问题上,与那些按正常的时间表拨打的人有所不同。例如,总的讲,那些在夜间工作的人在某些犯罪问题上给出的答案,可能与其他的被调查人不同。

第二，在可能的情况下，在反向电话簿上寻找电话号码[13]。第三，试着到当地的电话公司去找一些资料。遗憾的是，最后一种可选择的方法的实用性，在很大程度上取决于特定的电话公司对是否愿意帮助我们这一问题所持的态度。

移动电话、计算机和传真机的线路数量正在迅速增长。虽然大多数移动电话一般都被指派给了不同的交换站，而不是地上的线路，但这样的情况并非一成不变的。此外，有些被调查人可能会将电话从地上的线路转到手机上。在这样的时候，就技术和抽样框二者而言，所有我们能够做的是，训练调查员随时注意这样的情况发生的可能性。计算机和传真机线路比较容易识别和编码，就我们的目的而言，它们都属于非住宅号码。如果一个被调查人报告说某条线路既用于语音应答，也用于计算机或传真机，那么这个号码是否应该作为住宅号码，则要取决于被调查人所说的这一号码的主要用途是什么。

再次做拒访者的工作(reworking refusals)。在电话调查中的一个很重要的程序是拒访者转化(refusal conversion)。那些最初拒绝接受我们调查的被调查人，可能通过我们随后的工作而"转化"。尽管有关这一现象的研究不多，但是我们有理由认为转化之所以可能，其部分原因在于许多最初的拒访者开始都认为调查与自己无关。大家必须明白，我们的调查对我们自己是非常重要的，但对于一般的被调查人而言，它却意义不大。例如，有些人，在我们给他们打电话的时候不便接电话，他们会花一点时间与调查员约定再次打电话的时间，或会告诉调查员，请他们换一个时间再打过来；而有些人在遇到这样的情况时，则会拒绝接电话，但是在我们再次与他们进行联系时，其中有些人却转而同意接受我们的调查。专业的调查组织可以使用那些经过特别训练的调查员来实现最初的拒访者的转化，一般转化率可达到30% ~40%。虽然非专业的调查员的转化率可能没那么高，但是仍然还是会有一定数量的最初的拒访者可能转变态度，改而接受我们的调查。除降低拒访率之外，如果最初的拒访者在某些问题上不同于最初的合作者的话，这一程序可能还有降低偏倚的功效[14]。

一般讲，从最初的拒访到试图转化之间，需要有一个时间间隔。如果拒访是由于被调查人家中的"危机"所致，那么解决问题的时间可能需要几天。而许多被调查人甚至根本记不得我们曾经与他联系过。在遇有这样的情况时，我们大可不必提及上次的联系，最好还是从头开始。我们尝试使一个特定的拒访者转化所应采用的方法，在一定程度上取决于第一次联系究竟发生了什么。正因为如此，第一次联系的调查员必须记下拒访的具体原因，这一点非常重要。请大家记住，如果第一次电话联系还没有进展到户内随机选择的被调查人的程度，那么随后进行的联系有可能联系到另一个人[15]。当然，第一次拒访的这一情形应该是已经做过记录的。例如，假如一个人因马上要去上班而拒绝接受调查，那么将尝试转化的电话安排在与第一次试拨同一时间的做法

是再愚蠢不过的了，即使在那时打电话可以增加与被调查人联系上的可能性。同样，如果一个被调查人最初拒绝接受调查的原因是误将我们看作电话推销者，那么在做拒访转化时，我们便应早做准备，以便在随后的电话联系时，被调查人显露出这样的疑虑，很快就能把它们打消。

问题户(problem households)。有些住户，在与户内随机选取的被调查人取得联系之后，我们发现调查无法继续进行。有些被调查人根本不会讲英语，或虽然讲英语，但水平尚不足以接受调查。如果我们在一个人口中有相当多的不会说英语的人的地区(如迈阿密)进行调查，我们就有可能因为无法提供和操英语以外的被调查人，在目前的例子中，是操西班牙语的被调查人进行沟通的手段而产生严重的偏倚。然而在大多数规模不大的调查研究中，我们并不具备这样的能力。但是，这些住户仍然是定义总体的一部分，因而在计算回答率时，他们仍然应该被计为合格的住户。同样，那些病重到不能进行调查或者那些因有生理或精神缺陷而导致听力、理解或回答问题的障碍的人也都将在我们的调查中丢失。

最大程度地降低题项无回答

与单位无回答一样，没有回答的问题(题项无回答)的发生并不是随机的[16]。被调查人可能不愿意回答某些特殊的问题(如您是否携有防卫的武器这样的敏感题项)，或在回答另一些问题(如您认为明年家中有什么人有多大可能受到犯罪行为的伤害)时感到有困难。

第二类很有可能发生无回答的题项是那些难以回答或需要检查记录的问题。例如在一项医疗卫生方面的调查中，我们可能对被调查人去年一共在处方药品上花费了多少钱这样一个问题感兴趣。对许多被调查人来讲这一问题是很容易回答的，因为在这方面他们一分钱都没有花。但对于某些在这方面有很多花费的被调查人，这一问题就可能很难回答。在遇到诸如这样的问题的时候，在后一种被调查人中，有些人会试着回答或大致估计一下，而还有一些则很有可能简单地说一声想不起来了。有些题项，譬如“包括所有来源在内的全家总收入”，对某些被调查人来讲可能既比较敏感，也比较难以回答(如那些家中有很多收入来源的被调查人)。

同样，那些对问题、行为或事件需详加说明，但对被调查人来讲却是无关紧要的事实性问题，可能需要花费被调查人很多时间和精力来回答，其限度超出了许多被调查人所愿意付出的时间和精力的范围。虽然大多数人都知道在过去的 6 个月，他们看医生的次数，但是许多人可能记不住他们到杂货铺去了多少次。

许多诸如这样的问题的解决办法都有赖于问卷设计。在调查的后期，在调查员和被调查人之间已经彼此感到比较亲近之后再询问敏感的问题。在诸

如这样的问题前面，有必要再一次做出保密的承诺和/或说明问这一问题的必要性。对比较难以回答的数字性问题可以要求被调查人给出分类的答案，而不要要求他们给出精确的数字。

在试调查期间，我们应该注意是否有什么特定的题项的不知道或拒绝回答的比例异乎寻常的高。当然，某些题项出现这样的情况是由于逻辑跳跃模式所致，但其他类型的无回答，需要在试调查前和试调查中仔细地检查。

还有，编写质量很差的问题也会造成题项无回答。如果被调查人无法明白问题的含义，或必须费很大的劲才能理解，那么他就可能干脆不去回答它。下面将要讨论的调查员的行为也可能对题项的回答率有所影响。

调查员效应

在我们做到了在工具设计、事先通知和建立数据程序三者所能够做到的一切事情之后，我们的调查能否得到被调查人的合作这一问题便掌握在调查员手中了。不仅如此，调查员还可能影响回答某些特殊问题的态度和对这些问题给出的答案的质量。

我们不妨设想一下，在犯罪问题调查中，一个调查员与一户人家取得了联系，并随机地选取了一个成年人，而那个人却说："我真的没有时间，我马上就要出去，您为什么不去问我的妻子呢？何况她了解的新闻和政治要比我多得多。"调查员回答道："是吗，她可能的确喜欢我们的调查。"然后，那个人的妻子便过来接过电话，于是调查员就开始宣读调查介绍，而这位妇女却把他的话打断了，问道："为什么你们的大学要做一个有关犯罪问题的民意调查呢？""噢，"调查员回答道，"我不太清楚，也许这一调查与政府的改选有关。您知道，他们大概想表示他们的确非常想消灭街头的犯罪现象。"然后，调查员开始逐渐进入正题，问问卷上的问题："去年，您认为本州的暴力犯罪情况已经有所好转，还是更为严重，或没有什么变化？"被调查人答道："我女儿告诉我，她再也不在晚上到外面去了，因为隔壁的女孩就在家门口被人袭击了。"于是调查员说道："这就是说，您认为犯罪问题变得更严重了。"

这些交流过程的每一句话都显示出，调查员是如何在不经意间篡改了我们在调查设计和调查方法中精心编排的正确引导和指导调查的规定[17]。首先被调查人随机选择的规定被方便的选择所替代。原来规定的简介被缩短了，并插入了调查员个人的，且有可能对被调查人回答后面的问题有所影响的评论。最后，调查员自己从被调查人的评论中推断他对问题给出的答案，并将它强加给被调查人，而不是由调查人本人从与问题匹配的答案类别中探询答案。这一并非完全杜撰的例子告诉我们，调查员很容易，甚至是很自然地就会对调

查的质量产生影响。

在当面进行的调查中，调查员效应比电话调查更显著。调查员个人的身体特征都可以对被调查人的行为发生影响。例如，假如调查的内容与种族问题有涉，那么调查员的种族就可能会对调查有影响。调查员的面部表情和眼神，在电话上是看不到的，因而不会对调查有影响，但在当面进行的调查中，尤其是在进行与态度和敏感行为有关的当面进行的调查中，情况就不是这样了。就调查员问题而言，这两种数据收集模式之间的最大差别在于，二者实现监控和督导的可能性的程度是大不相同的。当面调查中调查员的监控和督导，无论是入户调查还是特殊人口总体的调查，都要比在集中的设备上进行电话调查的调查员更多地依赖于调查员自身。

我们将要花费大量的时间和精力，通过培训、监控来控制调查员的行为。此外，也是所有控制手段中最为重要的是，力求使调查员懂得自己那些漫不经心的、出于善意的行动对调查研究的成败是举足轻重的。我们必须记住，我们调查设计中那些最为关键的部分的实现必须依靠调查员。而我们的工作也常常都是从调查员的遴选开始的。

挑选调查员

不仅调查员的培训是很重要的，调查员的挑选同样也是很重要的。调查是一种技巧。它需要具备各种能力和条件，包括相当高的阅读能力、人际交往能力和善于说服人的能力等。我们要求调查员与素不相识的人接触联系，且在通常情况下，没有事先给出通知，要求人们配合我们的调查。不仅如此，调查员还要解释调查的目的，回答人们提出的有关调查或某些特定的题项的问题和根据调查手册的规定进行调查并精确地记录人们给出的答案。所有这一切都必须以相当专业的水准完成。有时还要根据实际情况相机处理。不言而喻，并非所有的人都能胜任调查所要求的全部工作。

有时，研究者在挑选调查员时是没有选择的余地的，例如在调查是一个课堂作业或雇佣一个调查机构进行数据收集时。不过，即使在调查是一个课堂作业时，也并非所有的学生都必须或适于做调查的。诚如我们所知，与调查过程中的其他方面的情况不同，实际的调查既需要了解调查过程所发生的问题，也需要掌握进行调查所需要的技巧。发现谁掌握了这样的技巧的最好方法是在培训之后做实地调查。一种可行的方法是让所有受过培训的人参加试调查，然后再确定什么人可以做调查，什么人可以做调查的监控，什么人可以做数据录入或担当其他工作。

无论我们是否真的需要聘用调查员，或只是对我们可资利用的调查员进行分配，我们都需要对下面几条导向性的意见加以考虑。首先，对预期的调查员进行结构性的测试是很有用的。结构性测试应该包括四个方面：理解和遵

守规定；争取被调查人的合作；恰当地朗读问题和正确地记录答案。对调查员掌握后两个方面的技巧的预期取决于他们事先经验的多少，当然，如果调查员以前有过类似的经验的话。但是，我们将会发现，即使是初涉调查的新手，在经过几次培训之后，无论在对态度比较勉强的被调查人做出反应的敏锐程度上，还是在自然地宣读调查问题方面都会有较大的进步。筛选可能的调查员的最好方法是手头握有几份具体的调查说明和一份问卷的草稿。我们给每一个调查员说明调查的目的，并告诉他们我们要求他们做什么。在允许调查员用几分钟时间阅读有关说明和问卷之后，由督导（或其他学生）扮演被调查人开始演练调查。先演练要求比较简单的调查及在回答问题时没有发生意外的问题。在随后几轮的演练中，“被调查人”的行为与理想的状况有所不同，人为地造成对调查员来讲是越来越困难的情景。观察者（或作为一个整体的一个班）将对每一个调查员打分。如果这样的测试可以进行，那么在两个相互分离的房间中的当事人双方，便可通过电话进行交谈，因而我们也就可以对调查员的调查技巧及在电话上的表达能力做出更加实际的判断[18]。

调查员培训

调查员的培训应该包括两部分：一部分是一般的调查技巧的培训，另一部分则是手头正在进行的调查的特定规定的培训（表 9.4）。一般技巧培训的时间的长短取决于调查员以前所有的经验，如果可能，这一培训最好由有经验的培训者担当。下面讨论的内容关系培训的最基本方面。它是对来自专业调查机构的培训手册的补充，应该严格遵守[19]。公立大学的调查中心自己编写的培训手册，对于其他（非营利）的研究者一般都是免费的或只收取少量的费用。

表 9.4　调查员一般培训要点

1. 争取得到合作
   - 确认或选择正确的被调查人
   - 介绍调查目的
   - 说服不愿接受调查的被调查人与我们合作
2. 进行调查
   - 开场白
   - 转入正题
   - 逐字宣读问题
   - 间接探问
   - 询问所有的问题并正确记录所有的答案
   - 遵照逻辑跳跃规则跳过应该掠过的问题
   - 正确记录答案：开放式问题

培训课程(和我们应该计划组织多次的培训课程)应该注重培养调查员的实际技巧。一个行之有效的培训计划,应该抓住培训课程中的每一个机会进行实际练习。我们一直都在强调一点,诸如我们在这一节的开头所阐述的那些行为,在交谈中貌似自然和出于善意,但对于实际调查是很不合适的。为了避免诸如这样的行为发生,我们不能只是简单地制定几条规定——当然,这也是必须的——但更重要的是,我们还必须要进一步告诉调查员,为什么这样的行为对我们的调查有决定性的影响。对实现这一想法很有帮助的一种方法是,要求调查员在培训课开始之前,阅读有关的背景资料。

实施调查

调查员培训中最为重要的内容是角色扮演,因而在培训过程中,有时需要进行模拟的调查。诚如我们所知,成功的调查并不需要理解过多的概念(尽管采取一些必要的措施,让调查员对概念有所理解是必不可少的),因为成功的调查需要的是某些争取得到被调查人合作和恰当地进行调查的技巧。而这样的技巧只有通过实践才能得到。一般讲,调查员对自己面临的任务的实践经验越少,他需要做的练习就越多。在培训上走捷径,表面上看似乎节约了资源,实际上却是对资源的最大浪费,因为它可能导致调查的失败,认识这一点也是十分重要的。如果调查员未曾在培训期间练习他们需要的技巧,那么他们就必定会在正式调查的时候,在真正的被调查人身上进行练习。显然,我们应该在正式调查之前,即在培训期间让调查员做必要的练习。

模拟调查的过程与被调查人筛选的过程颇为相似。在这一过程中调查员轮流扮演调查员和被调查人两种角色。每一轮实习都是高度有组织的集中演练一组特定的技巧。实习调查所需的各种技巧,应该一种接一种地进行,直至调查员能熟练、迅捷和妥当地处理遇到的每一种情况为止。记住,只要调查员已经在电话上正式开始调查,他们就不会再有时间去查阅说明、拿着电话犹豫不决或退回去从头再来。无论是被缺乏耐心的被调查人终止的调查,还是未被恰当地表达的问题都将导致低劣的测量结果。

各种练习的安排顺序,应以其在实际调查中使用的时间顺序一致。讲授每一阶段所需的技巧至少要安排一个练习。练习的数目和讲授每一个练习所需的时间取决于问卷的性质、调查员的背景和他们掌握调查所要求的技巧的速度。调查员应该了解,培训的最后一步将是对真正的被调查人做“活”的调查。用于这样的实习的个案应该选自研究的总体,而不是实际要调查的样本。当然,我们应该对那些最后一步实习的个案进行密切的监控。

对调查员最基本的要求可简单地表述为:完全按照书面的文字,照本宣科地逐字宣读每一个问题和过渡型陈述,包括每一类答案,不能有任何更动;即使在我们认为答案已经明白无遗时,一一询问每一个问题,遵守问卷指明的逻

辑跳跃模式;认真地记录答案,在遇有开放式问题时,答案要逐字记录。这些规则的表述十分容易,就如它们十分容易被违反一样。对于逐字宣读问题这一规则来讲,情况尤其如此。任何一个曾经做过调查的人,也许都会经历过一些似是而非的情况,即似乎必须加上一些评语或更动几个字才会使问题变得"清楚",或才会对被调查人有"帮助"。但我们必须再次指出的是,问题必须按原来书写的文本宣读,不能有任何修饰。虽然我们不能确定,对问题原文的每一次背离是否都会导致不可信的数据的产生,但是一旦我们为诸如这样的创造发明打开了方便之门,便会损害构成数据一致性的重要基础。我们必须努力设计出这样的问题,它们不会诱使调查员陷入修改问题的境地。

训练调查员掌握这种技巧的方法是,让每一个调查员轮流宣读问卷中的一个或一些问题,而让其他的调查员进行评议,评议的内容包括:是否逐字宣读、语速是否自然适当、在标点符号处是否有停顿和发音是否清楚等。调查员的语调应该与平常的谈话一样,不要像在念书。调查员常常想跳过这一实习阶段,因为他们自以为都是有文化的,都能足以胜任诸如这样的宣读。我们马上就可以给出许多证据,证明即使是那些受过良好教育的人,在用新问卷做第一个实习时,也常常会错误地宣读问题。使大多数听众感到语速过快,且有时还会读错。在这些错误没有导致在线的被调查人因恼怒而拒绝接受调查之前,它们似乎都是微不足道的。然而,正是这样的"微不足道",常常会导致被调查人挂断电话或对问题产生误解。处理调查问题不是一种朗读的技巧,而是一种执行的技巧。

在简单介绍了调查,并选取了被调查人之后,便是调查应该开始之时。我们应将问卷的格式设计成使它的第一个问题能使被调查人轻松地进入我们的调查,并能在大多数情况下使调查员都能顺利而平稳地进行调查。但是,诚如我们在前面的行为编码中所知,在调查员和被调查人之间进行的互动总会有这样或那样的问题。问题的起因概括起来有三种:问卷、调查员和被调查人。我们已经做过很多试验,证明通过问卷设计和训练调查员逐字宣读问题可以最大限度地缩小前两种起因。现在我们要讨论的是如何缩小第三种起因。

单位回答

调查员培训的两个关键的领域是争取得到被调查人的合作(单位回答)和执行调查(包括题项回答)。首先,我们先集中讨论争取合作的问题。诚如我们所知,近年来,在一般人口总体的调查中,得到被调查人的合作变得越来越困难。虽然训练调查员去争取得到被调查人合作的一般方法基本上没有什么变化,但是用于这一训练的实际练习时间却比以往更多,并已占据了培训时间中的大部分。

许多被调查人一般不会立即同意接受我们的调查。有时,他们希望对调

查有更多的了解,而不是仅限于调查简介所提供的那些内容。调查员必须随时准备好能迅速、一致和清楚地提供有关信息。一种使用比较普遍的实习方法是准备好一张"信息清单",上面有有关研究的组织者、调查的目的、在一般情况下做一个调查需要的时间和调查结果的用途的一个简短的描述。如果在试调查过程中或正式调查的初期,还有其他的被调查人经常提到关心的问题,那么这些内容也应该加到清单中去。提供一个被调查人可以用来核对调查的合法性的电话号码也是非常有用的[20]。少数被调查人可能希望再一次确认答案的保密性。但是对于大多数不愿接受调查的被调查人来讲,没有什么人会提出这个问题,他们只不过是对调查没有足够的兴趣,因而不愿为它花费自己的时间。

调查最主要的吸引力是它的主题——如果被调查人能得到那些信息。许多调查员填写的拒访都发生在调查主题提及之前。只要可能,简介应该直奔主题,而调查员也应训练成张口就谈主题——如果我们认为主题令人感兴趣或有什么特别的地方的话。当然,预达信件对这一问题也不无帮助。

不言而喻,不同的主题对被调查人的吸引力会有所不同。一般讲,对大多数可能的被调查人来说,犯罪问题调查比公共图书馆问题的调查更具吸引力。但是即使是令人感兴趣的、有报道价值的主题,仍会有相当数量的被调查人未必会立即同意接受我们的调查。调查员必须对这样的反应有所准备。在电话调查中,调查员可用于争取被调查人合作的时间是很有限的。记住,被调查人是很容易就会挂断电话的。

应对不愿接受调查的被调查人的方法有两种:①使被调查人有所事事和不断地说话;②专注于消除被调查人之所以不愿参加调查的特定原因。调查员必须倾听被调查人之所以不愿参加调查的原因是非常重要的,尽管在很多时候它与调查本身无关,只不过因为调查进行的时间不合适而已。如果被调查人或正要离家去上班,或正在电视上观看一场球赛,或正在家中处理个人事务,那么调查员马上就应有所警觉,立即放弃继续进行调查,并试着确定随后再次致电的合适时间。这种方法似乎属于常识的范围,在遇有这样的情况时,人们都会这样做,但是那些过分积极的调查员则有可能采取有悖常理的行动,在遇有这样的情形时,不暂时放弃调查或改变调查时间,便会导致被调查人勉强接受调查,最终被调查人从不愿接受调查变为拒绝接受调查。

即使调查员已经宣读了调查简介,告诉了被调查人调查的组织者和主题,仍然会有些被调查人对调查心存疑虑,怀疑它是伪装的电话推销或其他什么不正当的骗术。造成这一现象的部分原因是有些被调查人并没有认真听取调查简介。记住,他们并没有期盼打进来的电话,加之,许多市场营销公司的确常将自己伪装成调查。调查员必须对之有所准备,并快速地重复调查的组织者,如果组织者是一所大学或公信度很高的政府机构,则尤其要这样做。

许多被调查人将会简单地回答说,他们不感兴趣或没有时间。虽然这些回答听起来各不相同,但实际上指的都是同样的事情,即无非都是想赶快结束他们不愿接的电话。这时调查员应该做的最重要的事情是,必须想办法使被调查人有所事事,避免他立即挂断电话。如果调查的主题是颇受人们关注的,那么调查员就应该试着再重复一遍,且仍然要对自己听到被调查人的答话再次表示感谢。例如,调查员可以说:"一旦进入主题,大多数人都会发现自己乐于谈谈对犯罪问题的看法。我想,我们不会耽误您过多的时间。"

有些被调查人会说自己对调查的问题没有什么了解,特别是调查的主题听起来很有技术性,譬如与生物技术、某些环境或经济方面有关的问题。调查员需要使这样的被调查人确信调查并非知识考试[21]。这时,一个类似"我们只是想了解一下您的一般看法。这些问题并不需要有什么专业知识"这样的解释,可能就会使问题迎刃而解。

最后还需要加以说明的是,在有些时候,不愿意参加调查的并非实际的被调查人,而是户中接电话的人。在我们与某一住户联系,并用随机被调查人选择法选取被调查人的时候,样本户中的任何一个成年人都有可能成为回答问题的人。如果选取的被调查人并不是接电话的人,那么调查员应该要求和选取的被调查人谈话。这时,有可能会出现一个所谓的看门人,即那个接电话的但又不希望我们和被调查人谈话的人。其原因可能包括那些我们已经讨论过的和其他的原因。调查员应付看门人的办法,除了加上一些表示和被调查人简单地谈谈我们的调查愿望的言词之外,与应付不愿参加调查的被调查人并没有什么两样。如果在听取了我们的诉求之后,接电话的人仍然不愿意,那么我们应在另一个时间,再一次打这个电话,至少给自己一个越过看门人,直接与被调查人联系的机会。表 9.5 列出了调查员在遇有不愿作答的情形时,应该如何相机处理的一些建议。

题项回答

调查员的态度可以对人们回答问题的意愿发生影响。例如,一个有关受犯罪行为伤害的调查可能包括一个有关性侵犯的问题。如果调查员本身就觉得这个问题不太舒服,那么他自己对这一问题的态度,就可能对问这一问题的方式产生影响,也许他会把它念得飞快和/或把声音降得很低。诸如这样的行为势必使某些被调查人在回答这一问题时,比回答那些以同样的语速和语音宣读的问题感到更不自在。

未受过良好训练和严格督导的调查员的言词可能还会更加不恰当,以致对被调查人说诸如"如果您不想回答的话,您可以不回答"这样的话。诚如我们所知,在调查中的有些场合,如在开始调查敏感问题之前,或者在遇有被调查人不愿回答某一题项时,调查员应该提醒被调查人,他们给出的答案是会保

密的，这一点是非常重要的。我们应该让调查员对这样的场合有所了解。当然，我们的目的是在这样的情况发生时，能对之有所控制，以免某些调查员在遇有这样的场合时，鼓励被调查人跳过那些他们本来可以回答的问题。

表 9.5 如何应对不愿参加调查的被调查人：经常遇到的问题和补救方法

时间不当

- 不要试图强迫被调查人接受调查。
- 表示理解联系时间不合适。
- 判断是否要进一步确定再次打电话的时间。

怀疑伪装进行推销或有其他不正当的要求

- 重复调查的组织者和目的。
- 令被调查人相信调查并非电话推销或另有企图。如果可能的话，可提供一个 800 电话号码，或要求被调查人拨打接听方付费的电话，对调查情况进行核查。

没有时间/对主题不感兴趣

- 如果调查的主题是人们普遍所关心的，那么设法在这一问题上做文章，如果不是，则设法使被调查人了解，调查并不会花费很多时间。

被调查人对调查的主题不了解

- 将问题集中到有关观点的调查，不要理会知识方面的问题。
- 让被调查人了解我们的兴趣在于人们对问题的看法。

看门人

- 要求给被调查人一次听取有关调查是什么的机会。
- 再一次进行联系，尝试一下可否避开看门人。

在大多数调查中，题项无回答问题发生得不多。因为被调查人在同意接受我们的调查之后，一般都会回答我们提出的所有问题。调查员可以用于解决不愿参加调查问题的办法主要有两种。第一种办法是让被调查人了解为什么需要回答这个问题。例如有些被调查人可能希望了解他们为什么需要回答一些人口学的问题（如年龄、种族或收入）。在一般情况下，三言两语就能使这样的问题得到解答。如“我们需要将调查结果与有代表性的普查数据进行比较”这样的话。第二种办法是告诉被调查人他们给出的所有答案都是保密的，且不会与他们的姓名、地址或住宅电话联系起来。如果在使用了这些办法之后，被调查人仍然拒绝回答，那么最好跳过这些问题，继续去问下面的问题。过分地要求被调查人回答某一个特定的问题，有可能导致调查的中断和个案的丢失。有时，如果题项拒答发生在调查开始的阶段，那么有可能在以后再返回这一题项，那时被调查人就可能比较放松了，并对调查的合法性也不再

怀疑。

一个问题可能有答案，但却没有使用价值。这样的答案实际上与无回答无异。例如，一个记录得很差的开放式问题的答案（或在一个封闭式问题中记做“其他”这一选项的答案），我们可能无法了解它的确切含义，或无法对它进行编码。如果问卷中含有开放式问题，调查员必须逐字记录被调查人给出的答案。

最后，正像问卷的逻辑跳跃模式中的错误会导致某些被调查人跳过那些他们本来应该回答的问题一样，在用纸质的问卷进行调查的时候，调查员也可能在跳跃模式上犯错误。在培训期间的实习练习也同样会有助于调查员在实际调查过程中，正确地使用问卷进行调查。而 CASIC 系统的一个重要的优点正是它可以消除这一类型的无回答误差。

回答误差

我们知道，在被调查人误解了问题的含义、记不起有关信息或有其他回答问题方面的困难，或甚至有意提供假答案时，都会导致回答误差的发生。我们试图在工具的设计和测试期间发现并解决这些问题。

无论好坏，调查员都可能会对部分类型的回答误差有所影响。调查员可能会对被调查人如何理解问题的含义有影响，有时进而可能对被调查人给出的答案有影响。而调查员在这些问题上的效应最有可能发生在他们处理被调查人的有问题的行为时。

调查员必须学会处理的被调查人的行为主要有三种：①被调查人没用任何一类给出的答案回答问题；②被调查人不但不回答问题，反而提出问题；③被调查人的回答是有关题目的一个评论。在遇有这些情况时，调查员必须谨慎地加以处理，以免影响数据的结果或导致被调查人中止调查。

在遇有每一种这样的问题时，调查员都必须设法使被调查人能提供一种能编入给定的答案类别中的某一类合适的答案，当然，与此同时，还要注意和被调查人保持良好的关系。记住，被调查人无偿提供了他们的时间，因而我们应该力求对他们有所帮助。虽然调查员都知道自己得到的答案必须与封闭的答案类别相配，而无关的评论会费时间和金钱，但被调查人不一定会这样看，他们也许会把调查看作一种交谈式的交往，离题和跑题是不足为怪的。调查员的任务是使被调查人进入并保持主题——当然必须是十分巧妙的。要达到这一目的，最重要的问题是所有的调查员都要以相同的方法和方式来处理类似的被调查人问题，且这种方法方式不能影响被调查人对答案的选择。正因为如此，我们明确地告诉调查员，调查中什么样的行为是我们所期待的。不仅如此，我们还必须对此做出严格的规定，明确告诉调查员，在每一种情形下，什么样的行为才是许可的。

模拟调查中,在调查员实习应对每一种情形时,"被调查人"应该轮流做出各种不是一个好的被调查人所做的行为。在每一轮实习之后,整个小组应该对调查员的作为提出批评和建议。

在调查员没有得到可以接受的答案时,可以做一些探问,它是得到可以接受的答案的主要工具,即提出一个设计用来引出可以接受的答案的问题。调查员必须学会判别什么样的答案是合乎要求的,什么样的是不合乎要求的。而探问的使用是间接的,也就是说,我们并不直接建议被调查人,他们应该给出什么样的建议。

我们不妨来看一下犯罪调查的第一个问题:

一般讲,您认为,犯罪问题在您居住的地区很严重、不太严重,还是根本不严重?

1. 很严重　　　　2. 比较严重

3. 不太严重　　　　4. 根本不严重

如果被调查人的答案是"严重",那么调查员应该进一步探问,探问的方法是依次读出所有类别的答案。"您认为是……(宣读每类答案)?"调查员不应该从前面答案中推断答案类别的编码,即使是推断到非常相似的问题,也是不允许的。调查员的探问不应该做诸如"那么您认为是'很严重'了?"这样的推断。如果答案类别如上面所列,那么被调查人只是给出了部分答案。这样的答案将被调查人归入了一个答案类别的子集,这时我们的探问可针对这一子集中的被调查人。例如,如果规定的选项类别是"很满意"、"比较满意"、"不太满意"和"很不满意",而一个被调查人只是简单地回答"满意",那么恰当的探问可能应当是"您的意思是'很满意'还是'比较满意'?"不恰当的探问是在探问时没有重复所有有关的答案类别。在对每一个个案进行调查时,调查员都必须让被调查人自己选择答案的类别。即使同样的问题发生在同一份问卷中的不同题项上,我们也必须按照这样的方法,一个一个题项进行探问,不要删节任何有关的答案类别。尽管大多数被调查人很快就会挑出自己想要的答案,但总会有一些人需要在调查员对他们进行"培训"之后才能成为"好的被调查人",而这样的培训必须在调查开始的时候进行。被调查人一旦脱离了原定的轨道,再让他们回到原来的轨道是很困难的。

在培训阶段,调查员应该练习我们所建议的,也是他们将要在实际的调查中使用的探问方式。在练习怎样进行探问时,提醒大家注意那些颇为诱人,但却是不恰当的探问方式也是很有用的。例如,一个被调查人对上面那一个问题给出的答案是"严重",调查员不应该说:"那么,我是否可以认为您的答案是很严重?"我们再来看一下一个开放式问题:"您认为马里兰面对的最严重的问题是什么?"假如被调查人对这一问题给出的答案是"毒品"。一个不当的(有引导倾向的)探问会接着问道:"您为什么说是'毒品'? 难道是因为人

们在毒品问题上犯下了那么多的罪行?"一个恰当的探问应该是:"您能否更详细地介绍一下您的想法?"

我们必须处理的第二种类型的问题是由被调查人提出的问题。如果他们提出的问题与调查主题无关——例如一个被调查人问调查员自己有多么热爱调查工作,或自己对某一个具体问题有什么看法,调查员只要回答说,在调查期间调查员是不允许谈论自己的个人感受就可以了。如果被调查人谈的问题与调查有关,调查员必须学会不要妄加评论,除非对被调查人问及的问题有具体的规定。我们永远也不会了解,一句评语,即使是一个不经意的评论,也可能会对被调查人的答案有某种影响。

在对第三种问题——被调查人对调查的主题所做的评论——调查员应该不加评论,并应当引导被调查人返回调查本身。诚如我们在前面的例子中所知,在回答另一个犯罪调查的问题时,被调查人提到了自己的女儿曾经遭到过袭击,这时,调查员应该对被调查人提供的情况表示感谢,但不要做任何评论。为了达到这一目的,只要说一句"我知道了"也许就可以了。

表 9.6 提供了怎样处理调查中经常遇到的问题的例子。类似的角色扮演的练习也可用来练习这些例子,练习时可将表上列出的被调查人的行为打乱重排,以便使调查员在练习时能迅速地思考,并马上做出反应。记住,在所有这些情景中,所谓迅速而又正确的反应可能既意味着它能引导调查继续进行,

表 9.6 处理有问题的被调查人行为的常用方法

宣读问题原文的过程,因被调查人给出答案而被打断

- 调查员应该念完所有的问题。不过,如果它是一系列同一结构性题项的一部分的话,那么调查员应该根据自己的判断,是接受被调查人给出的答案,还是继续念完问题。

宣读问题答案类别的过程,因被调查人给出答案而被打断

- 如果问题与态度有关,调查员应该继续读完全部答案类别。
- 如果问题与事实有关,调查员应该接受被调查人给出的答案。

被调查人询问某一词汇或概念的含义

- 标准的答案是告诉被调查人"不论您怎么理解都可以。"对于比较特殊的术语,调查员可能应该提供统一的定义。
- 如果被调查人说不明白一个词或术语的含义,自己无法回答,调查员应该把这样的答案编码定为"不知道"。

被调查人询问更多的信息,或打听更多与问题无关的情形

- 调查员应该说:"您给出的答案只需以我刚才说的内容为依据就可以了,这样您认为……"并随后重复每类答案。
- 如果被调查人坚持在不了解相关情形或更多信息时无法回答,调查员应把该答案的编码定为"不知道"。

而不是中断，更不是拒绝；也意味着它能得到可靠的、无偏的数据，而不是错误的数据。

诚如我们所知，在整个数据处理过程中，随时随地都可能产生各种误差。一般讲，其中最严重的恐怕是单位无回答、被调查人误差和调查员效应。当然这些类型误差的大小主要取决于我们设计和执行的各个方面的质量。而经验告诉我们诸如这样一些类型的误差的大小，与数据收集的方法也有这样或那样的关系。表 9.7 显示了三种不同的数据收集方法的各种误差源的一般水平。

表 9.7 不同数据收集法的典型误差水平

	数据收集方法		
	电话调查	邮寄式调查	集体调查
单位无回答	低	高	低
调查员效应	中等	N/A	N/A
被调查人误差	低	中—高	中—高

质量控制

有若干种例行的程序可用于跟踪电话调查执行的质量。每一种都与我们刚才提到的、可能存在于大多数电话调查的各种重要误差源有关。其中有三种程序常用于跟踪调查员的执行情况：监控、回拨验证和数据检查。监控的方法是，在调查期间，在调查员和被调查人不知情的情况下，对每一个调查员调查的一小部分样本个案进行监听。这样的质量控制方法需要配备带有某种监听装置的中央电话设备，以便一个经过训练的督导员能在一般层次上（如和被调查人进行具有专业水准的交流）和非常具体的特定层次上（如逐字宣读所有的问题，在必要的时候对被调查人进行恰当的探问和恰当处理被调查人的问题或困难）对调查员的行为做出评价。监听的百分比取决于督导人员可用于监听的时间，但最多在 10% ~20%。我们建议在调查开始的阶段进行更多的监听，以便尽早确定和解决新的或经验较少的调查员身上存在的任何问题，同时也可尽早确定和解决所有在争取合作和调查进行中调查员遇到的各种问题。监听的关键在于立即反馈，无论反馈是正面的还是负面的，都要立即反馈给调查员。

在设备不允许进行监听的时候，我们可以采用回拨验证法对调查质量进行检查。在使用这种方法时，我们需要再次与每一个调查员调查过的个案中的一部分样本进行联系，以验证调查员是否做了调查，检查调查持续的时间，

并询问被调查人是否有什么困难或问题，如果有问题，还要了解调查员是如何解答的。究竟应该对多少个案进行回拨验证，没有什么硬性的规定，但一般大约在10%左右。不言而喻，就电话调查而言，监听法是最为基本的质量控制手段，应该成为电话调查首选的质量控制方法。

我们无法对面对面调查进行实时监控。那么一定数量的回拨验证是很必要的。虽然调查员弄虚作假并非调查中发生的主要问题，但是有时它也会发生。在调查实地，虚构一个调查（叫做非法的场外交易）的诱惑力要远远大于在装备中央电话设备的操作室的电话调查。可以记录调查时间长短及日期和星期几的CATI系统，给我们提供了对调查进行检查的手段，因而使弄虚作假更加困难。付酬的方式也会对调查员是否会弄虚作假有影响。如果调查员的酬劳是以完成调查的数量计算，那么它就会比按工作小时（不论调查员完成的问卷多少）计算的计酬方法诱使调查员更多地弄虚作假。

数据检查也是一种控制调查质量的方法。这种方法将在数据收集期间对收集到的数据制表和仔细检查。数据检查可以发现调查员调查过程中在问卷填写时发生的问题，也可以发现在早期没有发现的逻辑和问题设计上存在的错误。在数据检查时，我们将为每一个封闭式问题的答案制作成频数分布表，并为所有开放式问题逐字生成答案。然后再对这些数据进行仔细检查，检查的内容包括逻辑跳跃错误、丢失题项的模式或丢失题项数量是否过多、“其他”类使用是否恰当（即检查是否有本来应该归入某一封闭类别的答案，却被归入了“其他”）、相关题项间的一致性和开放式问题的答案记录是否清楚等。数据检查对我们是否能较早将数据中存在的各种错误检查出来，从而尽早采取纠正措施是很重要的。

自填式调查：邮寄式调查和互联网调查

尽管人们对互联网调查很有兴趣，且实际上也已经进行过了大量的互联网调查，但无论调查是用电子邮件进行的，还是在网站上进行的，它们还仍然处在发展的初期。我们还不能期待它们已经像其他数据收集的模式那样有了严格的操作规范和很高的调查质量。互联网调查的这一现状本身告诉我们，在我们决定使用互联网数据收集模式和选择调查程序时，必须要加倍小心。有关互联网调查研究的讨论超出了本书的范围。近来所做的大多数互联网调查，都是仿照邮寄式调查的模式进行的。我们不准备在这里对这一做法详加评论，只是希望读者能注意到这一点。有关互联网数据收集的主要信息来源之一系迪尔曼（Dillman ，2000）的有关论著，他所编写的邮寄式调查程序（Dillman, 1978），在过去的二十年间一直被人们视为这方面的范本。

有关互联网调查的各种观点和评论可以在库泊(Couper, 2000)的著作中找到。在一个有关邮寄式和互联网数据收集的导向性介绍中,我们将以常规的邮寄式调查程序为基础,并在适当的地方指出互联网调查通常与它有什么不同。这种比较保守的讲授方法,比较适合初涉调查研究业的新手。

电子邮件数据收集和网上数据收集

互联网调查可以用电子邮件进行,也可在网上进行。早期的互联网调查大多用电子邮件来收集数据。问卷或直接做在一个电子邮件中,或作为一个附件附在电子邮件后。被调查人在回答了问题之后,再将电子邮件或附件发回。现在我们仍然可以以这种方式收集数据,但是已经不多见了。我们之所以要在这里提到这一方式,是因为它的费用比较低廉,且所需的技巧也很有限,在有些情况下,仍然不失为一种有吸引力的、可供我们选择的方法。

电子邮件存在的三个问题使得研究者钟情于网上的调查。首先,电子邮件不仅系统种类繁多,且即使在一个单独的系统中,设置也各不相同,这就使得设计一份能送达每一个系统的问卷变得非常困难。例如,许多系统管理员都对进来的邮件容量设定了一定限度。在有些系统中,问卷可能根本无法直接进入,必须先被转换成一个附件(有些被调查人不知道如何处理这样的附件),或者干脆被删除了。许多人常常会自行将电子邮件从一个邮件地址转送到另外一个地址。如果他们在第二个地址填答问卷,那么返回的邮件将不再是研究者发去时的那个地址,从而使样本的追踪成为问题。

第二个问题是可供我们选择的、用于电子邮件的问卷设计法一般十分有限。如果我们将问卷直接写在邮件上,作为邮件内容的一部分,那么问卷只能是简单的文本,且被调查人必须将答案键入特定的括弧内。这就有可能增加被调查人误差的发生。逻辑跳跃模式也无法自动处理。附件固然可以以一种可执行文件的形式出现。但这样的文件不仅仍然可能存在诸如这样的问题,而且还有可能使这些问题变得更严重。

第三个问题仍然与附件有关。许多被调查人害怕打开他们不认识的个人和组织发来的附件,因为许多计算机病毒就是通过附件传播的。不言而喻,这样的担心势必导致回答率的下降。不过,对许多涉及某些总体(如高校的学生或某些会员团体)的小规模调查而言,电子邮件调查不仅费用低廉,而且效率也很高。我们已经指出了电子邮件调查可能会造成的某些类型的调查误差,除此之外,读者是否考虑到它还有可能造成哪些其他类型的调查误差?

网上调查的使用远比电子邮件调查普遍。正因为如此,以下有关互联网调查的讨论将只限于网上调查。在这样一些调查中,电子邮件用来与被调查

人进行联系，但是我们已经将调查的网址写入邮件，以便被调查人登录我们的网站填写问卷。有关互联网调查的讨论并未涉及特定的软件。与 CATI 系统一样，可供我们选择的调查系统不止一种。与其他软件系统一样，这些系统也随着时间的推移而不断变化，或退出市场。也与其他系统一样，在使用这些系统时，我们也常常会有各种焦虑和担忧，如系统的漏洞及缺陷、技术支持的程度和它必须与之交换的其他系统的兼容程度等。

单位回答

单位无回答是邮寄式调查中非抽样误差的最主要的来源。邮寄式调查的回答率一般低于调查员填写的一般人口总体的调查或许多特殊人口总体的调查。不仅如此，邮寄式调查的回答率甚至比互联网调查的还要低。无论是邮寄式调查，还是互联网调查，大量的补充联系是使回答率达到我们可以接受水平的主要手段。

这样一些填写方式的调查的回答率之所以比较低，其原因之一是这些方式本身就容易导致回答率下降。不言而喻。对于被调查人来讲，调查员本身就会给他带来某些社会压力，这种压力至少会使他去听取调查员请求他参与调查的呼吁。调查员可以根据被调查人的不同情况，调整自己呼吁合作的言词。而那些不使用调查员的调查则必须完全依靠文字或图形资料来说服被调查人参与调查，然后提供必要的指导，帮助他们填写问卷。我们事先很难知道我们设计的资料是不是恰当。与调查员操作的调查不同，常规的试调查无法在信件和问卷对答案有什么影响这一问题上给我们太多的信息。低质量的资料无疑会造成较低的回答率。实际上，低回答率本身就是一种反馈，而常规的试调查给我们提供的有关问卷修改方面的信息却很少。诚如我们所知，在常规测试中，焦点组访谈和认知方法对我们的帮助可能更大。

如果调查仅限于特殊人口总体，如某一组织的成员、某一行业或工作地点的雇员、高校的学生或其他相对比较同质的团体等，那么无回答问题可能便会大大降低。

与电话调查中调查主题的作用不同，邮寄式调查或互联网调查的调查主题可能会成为被调查人决定是否回答问题的主要因素。因为封面信件上列出的主题内容，被调查人可以根据信封上的回信地址猜出个大概，当然也可以浏览整个问卷，对它有更进一步的了解。一般讲，在对特殊人口总体进行调查时，如果主题特别适合他们，我们便可采用邮寄式调查。在这样的情况下，明智之举是尽可能地突出调查的主题。

而在互联网调查中，问卷是无法事先让被调查人知道的。如果工具比较简短，且似乎比较容易填写，那么被调查人便会注意到这一点，因而更有可能参与调查。在无法预览问卷的情况下，重要的问题是在封面信件中给出一些

有关问卷长度的提示。

事先通知将要进行调查

事先寄一封通知信或发一封电子通知邮件，告诉被调查人我们很快将要对他们进行调查，已经成为常规的邮寄式调查普遍采用的程序（Dillman, 2000）。它也是一种很有效的确定错误的通信地址的方法（通过邮局或电子邮件寄回或发回未曾送达的邮件）。如果我们有理由相信，我们的抽样框已经多少有一点过时，或含有其他什么误差，那么事先通知将会使我们对抽样框中的误差做出估计。它可以告诉我们，在邮寄问卷时，为了达到预定的被调查人的人数，我们邮寄的样本量将需要额外地增加多少。不仅如此，如果问卷的印刷费和邮寄费比较贵，那么去除错误的地址，可使我们节约一笔可观的费用。

在互联网调查中，我们应该事先发一封电子邮件通知。这样可以确认那些错误的地址。而它还有一个更为重要的作用，那就是又为我们提供了一次让被调查人知道一份合法的调查表将要送达的机会。现在，在网上散布的兜售信息和其他有害的电子邮件已成泛滥之势，这种现象势必会危及或已经危及互联网调查，我们不可等闲视之。

追踪联系

在事先通知之后接着再进行联系，那么为了完成一个调查，我们究竟应该进行多少次试联系呢？这一问题取决于我们手中握有的资源。即使我们无法告诉大家应该进行试联系的确切次数，但是我们可以肯定，无论是邮寄式调查，还是互联网调查，一定数量的试联系都是必需的。我们可以用下面介绍的备择方法为提高调查的质量做最后的努力。这一方法如果使用得当，便可以极大地提高被调查人的回答率。

在常规的邮寄式调查中，我们采用迪尔曼现在推荐的，应通过一等邮件和额外的特殊联，与被调查人联系四次（Dillman, 2000）。这些联系方式包括：

- 事先通知信
- 第一次寄问卷
- 明信片
- 第二次寄问卷
- 用特别的方法第三次寄问卷

在互联网调查中，我们也可采用与此相同的联系模式。当然，在互联网调查中，任何一份邮件都不会附有问卷。问卷一般总是发布在有关的网站上。然而，在常规的邮寄式调查中，每一个邮件（除了提醒明信片）都应该附有一份问卷。我们之所以要这样做是因为许多没有回信的主要原因只是因为问卷的误放、丢失或（无意中）丢弃（Dillman, 1978, 2000）。

在邮寄式调查和互联网调查中，在整个数据收集期间，各次邮寄的时间间隔要适当。后面邮寄的时间间隔要取决于前一次邮件的回信流量。一般要等到回信的流量成了涓滴一般几乎断流的时候，我们才考虑再一次邮寄。这样做有一定的优点（或说比较经济）。尽管互联网调查的回答模式有所不同，通常它都是非常迅速的，但是再次联系要等到回信量下降时再进行的逻辑却与邮寄式问卷是完全相同的。正因为如此，我们不仅必须跟踪我们收到了多少常规的回信，同时也要追踪在初始信件和此后的每次补充邮件发出之后，每天收到的回信的多少。

回答模式

一般讲，不论调查的总体是什么，只有很少的人会马上对一个邮寄式调查做出答复。之后，接着便会出现一个回信比较多的时期，在这一期间，每天收到的回信量几乎相同。提醒明信片在一定时间内，对保持这样的流量很有帮助，此后，随之而来的便是回信量的下降。在回信量下降期间的某一适当时候，我们应该发出第二封邮件，其中应附有一份问卷。为了阐明这样一种有典型性的模式，我们在图 9.1 中列出了全美县级官员调查的回信率。该调查的主题是如何确定县的优先项目。在互联网调查中，回答模式更集中在每次邮件发出之后的最初几天，那些打算回答的被调查人，通常几乎会在看到问卷之后马上就做出回答。

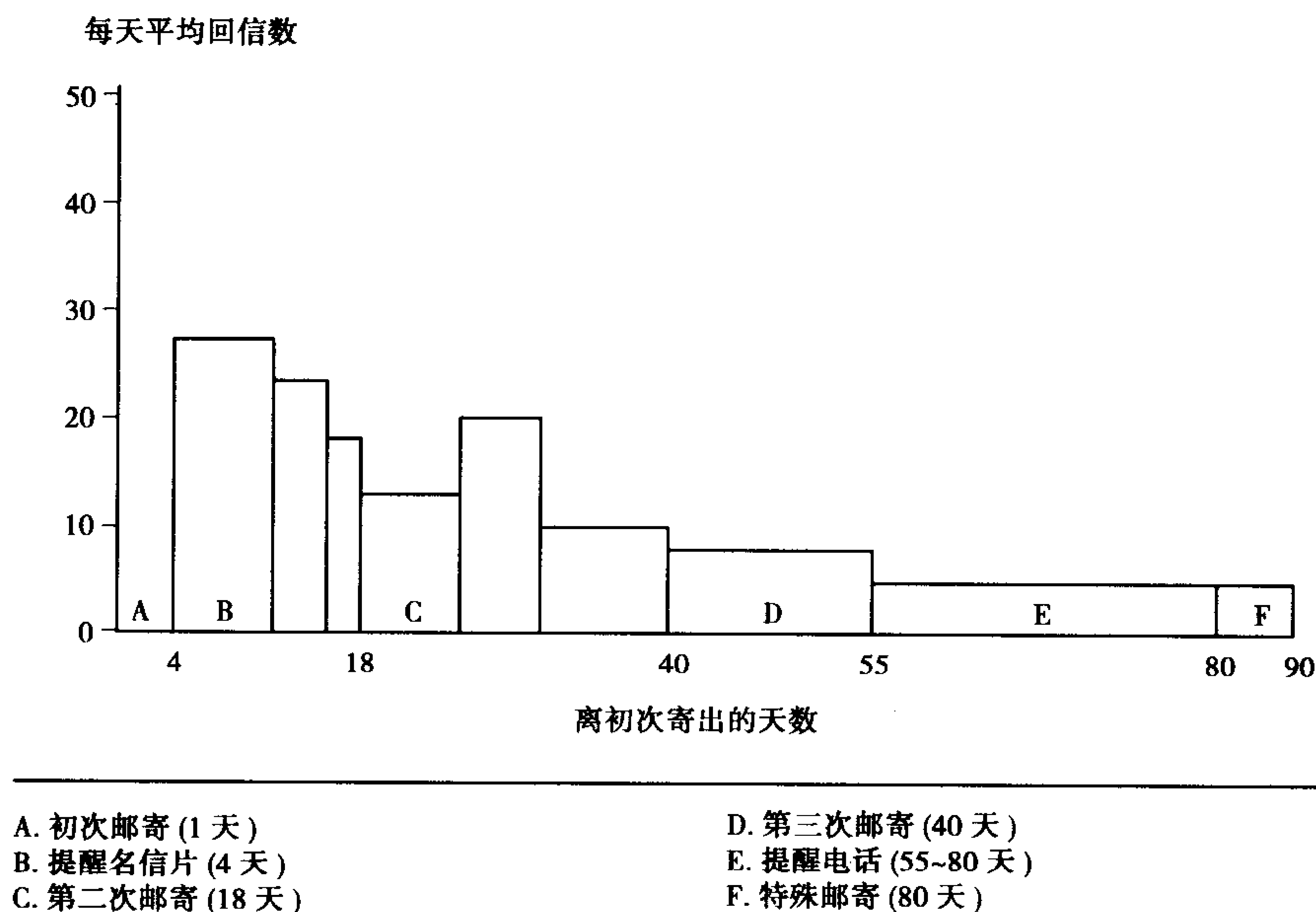


图 9.1 邮寄式调查的回信率

在邮寄式调查中，跟踪每天的回信数（或收到目标样本的百分数）是非常有用的。这种方法可以帮助我们发现回答率中存在的问题。图 9.1 显示，在初次发信之后，到开始收到回信之前，会有一个滞后期。在这一等待期，即滞后期，我们应该给所有的样本成员发一张提醒明信片。随后便会出现一个回信的大高潮，在几星期之后，这一高潮将逐渐消退。那时，我们应该考虑发出附有问卷的第二封邮件。此后的结果，与初次发出邮件之后颇为相似。之后，我们再对没有回信的被调查人通过电话进行联系，联系之后我们可能会收到不多的回信。而这也预示，我们的数据收集工作即将结束。

如果在对初次发出的邮件进行追踪之后，发现回信率未曾达到我们预期的水平，那么我们应该怎么办呢？首先，我们应该确定无回答是否存在某种模式。如果样本中的某些子群体比较不太愿意配合进行调查，那么我们也许应该把更多的资源投向这些群体。这时我们也许可以准备给他们再额外的寄一封信件，或安排给这些群体的无回答样本中更多的人发一封特殊邮件。如果我们的调查已经采用了某种物质奖励手段，那么适当增加这些群体的奖励金额，也许有助于这些群体的回答率的提高（Warriner, Goyder, Gjertsen, Hohner & McSpurren, 1996）。

无回答的原因

第二，试图确定无回答的原因几乎在任何情况下都是很有用处的。如果我们手头有无回答的被调查人的电话号码，那么我们就可以给他们打电话，搞清他们未曾将问卷寄回的原因。这样做也许对问题的解决会有所帮助。例如，如果我们在县级官员调查中发现，许多选出的县管理委员会的成员都没有立即收到问卷，因为他们中有许多人只是有一部分时间在委员会工作，而大部分时间都在从事其他的工作，那么我们也许应该将一些资源投入另外的渠道，通过电话得到这些人另外的通信地址。我们第二封信件，应该寄往新得到的备择地址。记住，在遇有这样的情况时，我们的抽样单位应当是人，而非地址。

我们可以从两个方面来考虑这些问题。首先，我们来考虑一下有关游艇主对于公共水道中废物的看法的区域性调查中的问题，然后再来考虑第二个调查中的问题，即有关学术调查组织使用的设计改进问卷的方法。如果我们发现许多游艇主在回答自己的游艇废物丢弃问题时，对自己给出的答案能否保密这一问题比较担心（因为他们知道某些丢弃行为是违法的），那么也许可以对第二封信件的封面信件做适当处理，专门提一下答案的保密问题。在调查组织的调查中，我们可能发现回答比较迟缓，因为我们要求的某些信息（如曾经使用过的独特的试调查方法的数目）并不是马上就可以提供的。那么随后的封面信件也许应该对此给被调查人带来的麻烦表示歉意和感谢，但可进而表示这些信息对提高回答率是非常必需的，因为调查的结果将在行业的研

讨会上发表。来自这些组织的被调查人,他们自己也经常会遇到这样的问题,而这就可能使他们理解我们的呼吁,尽力为我们提供所需要的信息。

在调查进行期间,虽然可供我们选择的方法是很有限的,但是有时只需要做一个小小的调整,其产生的效应便足以使整个回答率有很大提高,或足以使那些比较特殊的子群体的回答率有一定程度的提高,从而使我们能继续对这些群体做一些单独的分析。在使用这样或那样的程序时,关键在于,如果我们发现事情进展并不顺利,那么我们就应当采取一些必要的、可行的步骤,使不顺利的状况能有所改善。虽然事情的进展并非总如我们所愿,但一个优秀的研究者不会对此过多计较。

在互联网调查中,我们需要追踪的信息与邮寄式调查相同。当然,那时我们处理问卷的地点是在站点上(并非回信),追踪完成的问卷的数量工作由技术人员承担,这一点与邮寄式调查也有所不同。互联网调查的追踪与邮寄式调查的主要不同在于,它的追踪系统必须作为试调查的一部分预先进行检查调试。对这个问题人们在认识上存在一个误区,那就是误以为诸如这样的系统无论是在"箱子之外"的通用系统,还是专门为我们的研究所设计的专用系统,它都能按我们的期望,准确无误地运行。我在这里要再次提醒诸位的是,追踪系统必须在调查正式开始之前进行调试,而不在要调查进行期间才去寻找它存在的问题。

被调查人的选择

随机被调查人选择(在某些抽样单位内)在邮寄式调查或互联网调查中很少使用。这些调查方式一般不宜用于一般人口总体的调查,因此我们不需要讨论有关户内被调查人的选择问题。在大多数情况下,我们的抽样框都是一份列有合格的被调查人的清单。然而在某些企业或组织的调查中,我们可能只得到了一份职务清单,有时甚至情况比这个还要糟,我们必须将问卷寄给某个(不知姓名的)人,因为该人在组织中行使某种职责(如制定员工的规章制度),或有某种职务(如财务主管),或对组织有特别的了解(如了解用途各殊的具体开支)。我们距离指定有具体姓名的目标越远,问卷到达恰当的人选手中的可能性就越小,因而得到比较高的回答率的可能性也就越小。读者可以设想自己正在进行一项有关某一特殊类型的企业或组织的调查。考虑一下谁是那些我们希望填写问卷的人,这些人如何确定和如何找到。

与调查员实施的调查相比,在邮寄式调查和互联网调查中,我们丧失了很多控制的手段。在邮寄式调查和互联网调查中,我们不知道中介者是否是第一个看到邮件的人。而这样的情况发生与否,在很大程度上取决于我们所调查的总体的性质。在企业或组织调查中,常规的邮件到达目标被调查人手中之前也许已经做了充分的筛选,该人在组织中的地位越高,邮件被筛选的可能

性就越高。如果有某一备择的联系手段可行(如电话),它将对确定和说服中介者(如看门人)将问卷递交给被调查人很有用处。

我们有赖于某一个人来阅读调查说明,并希望他能按要求将问卷递交给应该填写的人。即使调查说明的文字简练,且被按要求执行,在许多时候(尤其是在对那些非常大的组织进行调查的时候),被调查人可能会要求他人代为填写问卷。这种做法将会导致什么样的误差可能要取决于我们的调查性质。

拒访转化

遗憾的是,许多可以用于电话调查和当面访谈的手段并不适用于邮寄式或互联网调查。很少有被调查人在回信中明确表示自己拒绝接受调查,并说明原因是什么。因此我们无法弄清楚没有回信的具体原因,从而也难以有针对性地在随后的补充联系信件中进行说服工作。我们不得不做一些虽然是一般性的,但却可能是多次的呼吁,请求被调查人的合作。例如一份附在补充联系的邮件中的封面信函,便可以提及诸如被调查人想必非常忙,或将第一份问卷放错了地方(或无法进入网站)等这样的事情。但是被调查人之所以没有回信的原因可能与二者都无关。在邮寄式调查中,我们还可以试一试另外的方法,那就是专递邮件和/或要求核对通信地址。这两种方法可能都比较有效。核对通信地址的要求要提前实施,以便随后发出的邮件能到达正确的地址。专递或其他专用方式,因费用比较高,只能在调查的后期,集中用于少量的个案。在使用这些形式的邮递时,最重要的问题是要事先确定,无论被调查人是否在家,邮件都要保证能被送达。这样的邮件必须直接送到,而不需要到邮局领取,或还需要其他什么手续,否则这样的邮件也许会适得其反,不但不能提高回答率,反而使被调查人更加感到不快(关于在一般人口总体的邮寄式调查中,如何处理这类问题的讨论,请参见 Dillman, 1978)。

无回答被调查人样本

最后,我们来考虑一下在单位无回答比较严重的情况下的无回答的被调查人样本问题。在邮寄式调查中,无回答的被调查人通常被用来评估它与回答的被调查人样本之间存在的差异。此外,它也被用来提高调查的总回答率。在采用这一方法时,我们首先要从一个未曾将信寄回的被调查人的样本中选取少数成员,然后尝试通过其他办法,大多数情况下都是通过电话,对他们进行调查。然后将这些人的特征与那些已经收到他们的回信的被调查人做比较,以对邮寄样本的可能偏倚的方向和大小做出评估。虽然以这样的评估结果做的统计学的修正已经超出了本书讲授的范围,但是能对这些“不愿参加调查的被调查人”很可能都是来自某一特定性别或背景或都是不太同意问卷

提出的某些建议的那些人这一点有所了解,是非常有用的。不言而喻,无回答的被调查人抽样的方案必须事先制订,它也是整个设计方案的一个组成部分。但是它的具体实施步骤的确定则可以在我们对无回答的性质有了更多的了解之后。

题项无回答

在邮寄式调查中,丢失数据或模糊不清的数据的处理问题是一个比较重要的问题。在互联网调查中,我们可以对问卷编程,使被调查人在想跳过某一问题时,无法向前移动光标。但是这种做法实际上有喜有忧,好坏参半。一方面,它的确可以降低题项无回答,但另一方面,一个本来非常想跳过某一问题的被调查人,因无法跳过而不得不键入一个假答案,或者甚至因此而拒绝完成整个问卷。这里我们还会遇到一个道德问题。因为通常我们都会对被调查人说,他们可以不回答任何他们不愿回答的问题。因此我们必须认真地考虑这一问题,那就是我们是否仅仅因为技术上的可能,便置我们自己做出的道德承诺于不顾。

在收到一份带有未回答题项的问卷时,我们可有三种选择:忽略这些题项,把它们的编码在数据集中定为"丢失";尝试确定这些答案"应该是"什么,并将确定的值键入;再次与被调查人联系,得到丢失的答案。

在决定采用其中哪一种方法时,我们必须弄清哪些变量是我们研究的关键。显然,所有的变量都是有价值的,否则我们就不会将它们列入问卷,但是这些变量中毕竟有些变量更为重要。例如,如果我们的分析主要涉及种族差别问题,那么种族数据的丢失,无疑会使该个案对研究的用处不大。另一方面,如果种族差异并非我们研究的重点,那么这一变量丢失了,我们大可不必为此而过于担心。

如果丢失的数据数量不大,且涉及的题项也非研究的关键,那么我们就考虑选择那种花费较少和不花费资源的方法。我们既可以忽略不计,也可以通过跨变量的分析,确定丢失题项的可能答案。例如,如果某一个被调查人的问卷性别项没有答案,但是在问卷其他题项的问题中,该被调查人报告自己就读于一个教会女子中学,那么我们就可以在丢失项中键入与此项答案一致的答案"女"。在使用这一方法时,我们必须非常小心。诸如这样的归因推断导致的误差有可能使数据质量变得更差,而非更好。

在数据丢失可能是由于错误地理解"跳跃"的说明而引起,且该题项又对研究十分重要的时候,最为恰当的处理办法,无疑是再次与被调查人联系,以得到正确的答案。如果我们可以通过电话很快便能与被调查人再次取得联系,那么这一方法便更加有吸引力。如果电话是长途电话,那么费用问题就必须加以考虑。所以事先留出一小笔资源——资金和时间表——以备这不时之

需的做法不失为一种明智之举。

追　踪

邮寄式调查的追踪程序是相当简单的(参见 Mangione, 1995; Dillman, 2000)。我们已经指出,追踪邮件的回答情况必须按回信的日子进行,以便为进一步的补充联系做计划。这样一种追踪只要用电子表格和记录个案编号的简单系统就可以进行。在互联网调查中,追踪程序可能比较复杂,但是究竟能达到什么程度,则在很大程度上取决于我们使用的应用软件的性能。虽然讨论备择的互联网调查的软件超出了本书的范围,但是我们还是要提醒那些初涉互联网调查的研究者,在确定应用软件的使用时,必须兼顾软件在问卷管理和追踪两个方面具有的兼容性和操作的灵活性。

注　释

1 虽然出版年代略嫌久远,但却是最为全面的有关误差处理方面的专著是 R. 格罗夫斯(R. Groves,1989)的《调查误差和调查费用》(*Survey Errors and Survey Costs*)。

2 注意,误差一词并非专指错误本身,它的含义要广泛得多,广泛到甚至把诸如将来自邮寄的问卷的数据错误地键入计算机这样的错误也包括在内的程度。

3 量度过程本身也同样会产生误差。当然,诸如这样的量度误差不仅仅限于调查的领域,而是遍及整个经验科学。

4 注意,对于将要产生的假报告,如引起人们把年龄报小,或以某种特定的方式回答某些其他的问题等这样一类偏向,我们无计可施。与日常的用法相反,调查研究中的偏倚是指效应,而非意图。

5 这是我们之所以要把标准化地收集人口资料作为调查的一部分的原因之一。它使我们得以将样本的人口学分布与调查地区中满足目标总体定义的所有的人口的分布进行比较。

6 为了在一定程度上解决这一问题,在数据收集完毕之后,我们要进行一些加权修正。然而,正如我们在第 10 章要看到的那样,加权会使调查数据的分析变得复杂,从而增加其他的误差源。

7 有些被调查人拒绝接受调查的态度比其他人强硬。那些不太强硬的拒绝可成为软拒绝。

8 虽然这一名词最近才被引进。实际上,它已经发展到遍及调查研究中计算机使用的所有方面,如传送、处理、分析和分发等方面。

9 在调查作为研究课程教授内容的一部分时,需要使每个人在编码、监控或其他非调查性工作中不断进行变换。

10 回答率的定义是调查的户数除以合格的住户数。合作率的定义是调查户数除以调查户数、部分完成调查的户数和拒绝接受调查的户数三者的和。在调查结束的时候,回答率是数据收集是否成功的主要量度指标。在数据收集期间,是测量调查进行是否顺利的比较好的指标。这是为什么?

11 对于由组织实施的调查而言,实际上这是一种常规,而不是例外。在计划准备这样一种

类型的调查时,即使只是打了一两个约定调查时间的电话,也应该给对方一定的费用补贴。

12 对于一般人口总体的调查而言,大约80% ~90%的电话应该在平日(周一至周五)下午的6点到9点、星期六和星期日早上晚一点的时间和下午以及星期日的晚上打。在其他时间打的电话,目的主要是为了去掉营业号码和与工作时间异常的被调查人进行联系。

13 反向电话簿是一种按电话号码的次序排列的,而非按用户的姓氏笔画排列的电话簿。海因斯和R. H. 多内尔公司是这样的电话簿的主要出版商。

14 在对来自若干调查的拒访转换的被调查人和其他的被调查人进行比较之后,布莱尔和瞿(Blair and Chun , 1992)发现,在"不知道"的个案数和题项无回答的数量上,在最初的合作者和经转换的合作者之间是有所不同的。

15 户中的随机的被调查人一经选出,便应留作调查的目标回答人,不论以后的试拨发生的情况是什么。

16 有一种例外发生在被调查人仅仅是因为疏忽而漏答了邮寄式调查的某一题项时。例如,在回答标有答案类型适用于系列中所有的问题时,譬如每一个题项的答案选项都是"是"和"否",被调查人漏答了其中的一个或几个题项(Rasinski, Mingay, & Bradburn, 1994)。

17 我们也应该意识到调查员和被调查人之间的交流,也有可能会影响到被调查人随后的举止行为。

18 如果这样一种类型的培训是组织实施得当的,那么它就应该在一或两个房间内安置上一个喇叭。

19 密歇根大学调查研究中心编著的《调查技术概论》(*General Interviewing Techniques*)(Guensel, Berckmans, & Cannell, 1983)是一本上佳之作。

20 如果调查不是在当地进行,那么首选的号码应该在1 ~800之间。对不论什么范围的调查来讲,在规定的工作时间,这个号码都应该是有专门的人员来接听的。

21 重要的问题在于,一般讲,调查,特别是那些似乎令人望而生畏的调查,一定不要从那些知识性的问题开始。记住,所有的被调查人,特别是那些本来就不太愿意回答我们问题的被调查人,如果最初的几个问题比较困难,或令他们觉得不太舒服,他们仍然有可能改而选择终止我们的调查。

第10章 若干比较重要的专题

Special Topics

本书涉及的某些专题并非调查设计和执行的组成部分，然而它们的重要性却不容置疑。我们非常强调资源配置的重要性，这一观点贯穿了本书的始终。了解调查的预算是如何确定的是了解资源使用的基础。数据收集涉及与可能的被调查人的互动。正是这些被调查人，应我们的要求将自己的时间和资源贡献给了我们的研究。在如何恰当地对待研究对象的问题上，有些与道德有关。我们最后要告诉读者的是，有关调查的方法学的阐述是研究文档的一个不可或缺的组成部分。

调查研究中的道德问题

在前面的各个章节中，我们始终都把注意力放在了职责问题上，我们自己主动假定，我们将尽我们所能把我们的研究工作做得最好。但是一个严肃的研究者的责任并不仅仅限于研究工作的技术和理论方面。我们同样也对那些同意参与我们研究的被调查人、我们的同道和我们的研究结果的使用者负有责任。虽然对这些问题追本溯源地进行深入地讨论，超出了本书讲授的范围，但是对于初涉调查研究工作的新手来讲，能在一般的意义上，对这些问题有一个最低限度的了解则是十分必要的。

在正确对待被调查人的问题上，我们必须注意两个核心概念：知情同意（informed consent）和保密。在一个调查中，虽然为了得到被调查人的合作我们尽了很大的努力，但是被调查人同意接受我们的调查的前提必须是对调查已经知情。这就意味着我们一定不要在研究的性质和目的问题上误导被调查人。我们必须如实回答他们提出的有关项目的问题。这样的问题包括项目的组织者是谁、研究的主要目的、调查要花费多少时间和精力、调查主题的一般性质及数据的用途等。我们既不能诱使或迫使被调查人参与我们的调查，也不能在他们同意参加我们的调查之后，诱使或迫使他们回答某些特别的问题。

在被调查人同意参加我们的调查之后，我们便有了对他们的答案保密的义务。无论我们是否已经明确告诉被调查人，实际的情况就是如此。调查结果或数据集中可以对被调查人的身份一一加以确认的信息决不能外泄。

这些道德规范已为专业的调查研究组织所认可，并受到大学或其他从事人口研究的组织的人类主题评论委员会（human subjects review committees）的专门检查。与一个由单独的研究者个人执行的项目一样，由一个研究团队从事的调查项目，这些义务也同样适用。实际上，在集体进行的项目中，我们需要注意承担的义务可能更多，因为在许多人都对样本和数据有利害关系时，在不经意间违反了这些道德规范的可能性必然会更大。

将被调查人的答案在研究团队之外公布或议论是不合适的。而在被调查人最初接受我们的调查的目的之外，再次与他们进行联系也同样是不合适的。用于调查的样本清单，在未曾再次取得被调查人同意之前不应当提供给他人（即使是其他合法的研究者）。如果要将数据提供给其他方面使用，应该去掉那些可以将答案与被调查人个人联系起来的标识性信息。

作为研究者，为了保证我们的研究结果的信度和效度，我们都必须付出一定的努力，采用适当的方法收集数据，并认真的解释得到的数据。而在我们的调查报告和讨论中描述和展示这些方法则是我们不容推辞的责任。当然全面而清晰地阐述调查的方法学的理由还不止这一点。我们将在本章随后几节中对这一问题进行详细的讨论。

方法论报告

在撰写用于调查设计和实施的方法报告时，我们将从另外一个角度审视工具设计、抽样和数据收集中的若干技术问题。在方法论报告中，调查的每一个部分的误差性质都是有关调查结果的信度和效度的信息。之前已经提到过的几个问题，将在方法论报告中重提，但更多的是从使用者的角度。

我们力求通过一系列的设计和实施的决策，最大限度地提高调查结果的信度和效度。为了对调查结果做出公正性评估，读者（或数据集的用户）不仅必须知道我们关键的决策是什么，同时还必须知道我们曾对调查的结果做了一些什么决策。研究者在忙于转向实质性结果的分析之前，常常省略了调查的最后一步，或准备了一个报告，但却是不完整的。在这一节，我们将通过若干例子来说明方法论报告的某些性质。这些性质都会对调查结果及其阐释有举足轻重的影响。进而，我们还将对一个典型的方法论报告的主要组成部分做一些讨论。

方法论报告的用途

对调查的方法论上的程序和结果进行正式的文字阐述是很困难的，但是用具体的例子来对它们加以描述却是比较简单易行的。我们不妨来思考一下来自犯罪问题研究假设的发现：

去年是否出于保护自己的目的购买过枪支或其他武器？

答案	分布
否	66%
是	23%
不知道/拒绝回答	11%

如果我们只是简单地报告 23% 的被调查人回答“是”，那么我们得到的只是画面的一部分。如果我们得悉调查的回答率是 45%，那么这一数字将会对这一估计数的置信度产生什么样的影响呢？如果回答率是 78%，影响又会是什么样的呢？不言而喻，我们凭直觉就能意识到回答率越高，我们的发现得到的支持就越强。但是从高于直觉的角度看，高回答率之所以很重要是因为我们可以较少地为这样一个问题而担心——假如那些未回答问题的被调查人接受了我们的调查，那么调查的结果是否有可能因此而有所变化？记住，我们会因为用选取的概率样本来映射目标总体而已经有了一种统计偏倚。然而，假如我们没有能调查到所有的样本成员，那么我们又会使完成的样本和选取的样本有差异，从而使我们映射总体的能力有所减弱。因为调查很少有，或者说只能在一个特定的环境中才能有 100% 的回答率，无回答的可能影响实际上是我们永远也无法避免的，因此我们必须弄清那些未回答问题的被调查人与回答了问题的被调查人究竟在什么地方有所不同。

当然，我们无法确定无回答影响的确切程度，但是这种影响是与无回答的数量成比例的[1]。如果无回答率比较低，那么那些我们没有能调查成功的人，必须与被调查人有很大的不同才能使最终调查的结果有所变化。相反，如果我们没有能对相当大的一部分人进行调查，那么最终的调查结果很容易会因此而发生变化。

为了对无回答问题的可能效应做出判断，我们也必须对样本的分布加以考虑。如果在一个是/否题项上，我们得到的样本估计值为 93% 的“是”和 7% 的“否”，那么我们就不必过于担心可能的无回答误差对这一估计值是否可信的影响。相反，如果我们得到的样本估计值是 55% 的“是”和 45% 的“否”，那么我们就需要担心可能的无回答误差对这一估计值的可信性的影响。在前一种情况中，即使未回答问题的被调查人明显不同于被调查人，总体中的绝大多数人答案类别仍然可能为“是”。而在样本中“是”和“否”两类答案的人数十分接近时，如 55% 和 45%，情况就不是这样了，这时，如果未回答问题的被调

查人分布与被调查人明显不同,那么最终的发现就会有一定变化。

在这同一个有关犯罪问题的例子中,我们再来考虑一下两个假设的结果。如果样本中妇女的百分比是64%,而在成年的总体人口中,妇女的百分比只有53%,那么结果会怎么样呢?或在拒绝回答这一问题的被调查人的百分比为11%时,结果又会是怎么样的呢?如果与男性相比,更多的女性会出于自卫的目的而携带武器,那么我们就有理由考虑是否需要对这一数据用性别加权(对数据进行修正,以使它能与总体的分布匹配),因为未加权的估计值有可能过高地估计总人口中已经购买了武器进行自卫的人数。

最后,我们来考虑一下题项拒答的问题。可以想象,那些拒绝回答这一问题的人,其已经购买武器的可能高于那些回答了这一问题的人。可以令我们认为事实的确如此的理由有两个:第一,许多人可能认为购买武器是一个私人的,且多少有些敏感的问题,因此不愿意回答这一问题。第二,那些虽然购买了武器,但是尚未得到许可证,或没有遵从其他法律规定的人可能不愿承认自己已经购买了武器。考虑到这两个原因,我们有理由认为23%的“是”可能是一个低估的数字。虽然在缺乏有关的补充信息的情况下,我们难以对得到的23%这一数字做出正确的判断,但是我们现在可以知道,样本统计值本身,离开了它的方法论依据,它所提供的信息有时是非常不确定的。

在许多时候,尽管我们已经尽了最大的努力,但是我们的样本的人口学特征的分布还是与我们用作权数对样本统计值进行修正的总体有很大的不同。加权法的使用不仅是非常普遍的,而且它对结果所具有的效应也是很大的。我们可以用加权法对数据进行进一步的处理。

我们假定在全州完成了800个调查,其中560(70%)个被调查人来自城市,而其余240(30%)来自农村。我们再进一步假定二者在是/否问题上的差别,在城市被调查人中,是和否的比例分别为40%和60%,而农村则为87%和13%。二者的联合分布已在表10.1中列出,为54%和46%。

假定在查阅了美国近年的普查数据后,我们发现该州85%的成人居住在城市地区。这样在我们的样本中,这一部分人显然被严重低估,不言而喻,农村地区的人数则因此而被大大高估。这种相当不准确的代表性无疑将对我们有关州的发现产生影响。我们通过加权对城乡的分布比例进行了修正,使之与普查的数据相配。我们采用对城市被调查人计数略多于1,而对农村的被调查人的计数略小于1的方法来对估计值进行修正。在这一例子中,权数只不过是一个普查百分数与我们样本的百分数之比[2]。加权以后的结果显示在图表10.1的下半部分。这样,我们就有了加权的样本估计值,是和否的百分比分别为47%和53%,它几乎刚好与未加权的数字相反。

表 10.1　加权和未加权数据比较

	未加权估计值		
	城市	农村	总计
是	224(0.4)	209(0.87)	433 (0.54)
否	336(0.6)	31(0.13)	367 (0.46)
总计	560	240	800
百分比	70%	30%	100%
	加权估计值		
	城市	农村	总计
是	272(0.4)	104(0.87)	376(0.47)
否	408(0.6)	16(0.13)	424 (0.53)
总计	680	120	800
百分比	85%	15%	100%

这一似乎有些极端，但并非完全不真实的例子告诉了我们有关加权的几个非常重要的问题。任何人在撰写方法论报告(或分析调查数据)时，都应该清楚这样几点：第一，权数可以对结果的估计有非同一般的影响。它们经常为样本设计所要求(诚如我们在有关抽样问题讨论的第 8 章所述)，或如这一例子所示，为处理在无回答问题上存在的差异所需。它们既不能随意使用，也不能以研究者的偏好为转移。相反，我们必须审慎地对它们加以选择和恰当地予以使用[3]。

第二，只有在被调查人和未回答问题的被调查人之间不存在明显的差异时，加权才能有效地处理无回答偏倚。我们之所以要对城市被调查人加权，是因为他们的回答率低于农村，因而他们在样本中的代表人数过少。但这样的加权，只有当城市中的被调查人在实质性题项上与城市的未回答问题的被调查人答案大致相同时，才有望获得成功。为了阐明这一点，我们将一个相当大的权，120 个个案加到了城市样本上，实际上，这就意味着如果要使我们的样本能精确地反映城乡居民的比例，我们必须增加 120 个城市的样本。更确切的讲，那些“丢失”的被调查人将以与那些已经回答了问题的被调查人同样的方式，来回答我们的问题。这就是说，要使加权的是/否分布和未加权的保持一致。

实际上，我们并不知道加权在处理无回答偏倚问题上的实际效应。有可能它只有部分效应。鉴于这样的不确定性，我们应该在方法论报告中对有差别的无回答问题展开认真的讨论(特别是在情况如本例这样严重的时候)。

最后，我们还应该了解，只有在被加权的组别之间存在着差异，权数才会对总估计值有影响。如果是城乡居民的是/否分布没有差异，那么加权将不会在这一特定题项上对全州的结果发生影响。但是，无独有偶，我们必须记住，调查是由许多变量组成的，加权虽然对某些变量没有影响，但对另外一些可能

会有影响。

虽然加权是一种很重要的方法,但是我们在使用时必须十分小心,因为我们在对一个变量加权的同时,很可能对数据集中的其他变量有影响或做了调整。这就是说,如果我们对城乡比例做了加权,那么我们也可能对性别、收入或其他变量做了调整——在一定程度上也为加权的变量所纠正。种族和收入可能存在明显的城乡差别。在我们对城乡比例进行加权的时候,我们也会对这两个变量产生影响。如果我们把城市居民比例加大,而城市中非白种的居民更多,那么在我们总样本中,非白种人数的比例也会因此而上升。另一方面,如果城乡居民的性别比例不存在差异,那么对城乡居民比例这一加权,将不会对总样本的性别比例发生影响。

我们之所以要提供详细的方法论报告的另外一个原因是,它将有助于与其他研究做比较。让我们再一次回到武器问题,例如,如果我们分别进行了两个内容相同的调查,但得到的估计值却明显不同。这在很大程度上可能是由抽样误差造成的。当然人们可能会问,这样说的根据是什么。如果马里兰犯罪问题调查,在对 800 个被调查人的数据进行了分析之后,发现 23% 的被调查人已经购买了武器,而这一估计值的标准误差在 95% 的置信水平上约为 2.9% 。如果我们将同一问题询问 700 个弗吉尼亚的被调查人,得到的结果中“是”的比例为 40% ,而在相同的置信水平上,标准误差为 3.6% 。那么只考虑抽样误差,马里兰的值域是 20.1% ~25.9% ,而弗吉尼亚则是36.4% ~43.6% 。显然,马里兰的最高估计值也远远低于弗吉尼亚的最低估计值[4]。

当然,这不同的发现可能反映了两个被调查的州之间存在的差异。弗吉尼亚州居民购买武器的人数就是比马里兰多。但是诸如这样的有差异的结果也可能是因为调查方法使用不当造成的。

在试图理解这些差异的过程中,我们首先来看一下两个调查的主要方面。如果数据收集方法不尽相同,那么差异就可能由此而产生。较之电话调查和面对面访谈,邮寄式调查的被调查人可能更愿意承认自己拥有某种武器。题项在问卷中编排的位置可能也会对答案有影响。如果一个调查,在调查的合法性尚未得到确认及调查员和被调查人之间尚未建立起比较融洽的关系之前,很早就问到了这个题项,那么它在这一题项上得到的回答率就可能低于另一个在调查的合法性已经得到确认,且调查员和被调查人之间的关系已经比较融洽之后,才问及这一问题的调查。最后,如果两个调查是由两个不同的组织(或研究团队)实施的,那么这一差异就有可能是“机构效应(house effect)”的结果。所谓机构效应是指机构在培训时告诉调查员,他们在进行调查时一般应该采用的方法,如被调查人在给出的答案为“不知道”时,需要还是不需要做进一步的探问,对结果可能产生的影响。

例如,如果在马里兰调查中,这一问题问得比较早,那么就可能产生某种

联合效应，导致只有较少的被调查人报告自己购买了武器。如果弗吉尼亚调查采用邮寄式，而不是像马里兰那样采用电话调查，那么这一差别就有可能加到效应上。在这一例子中，两种效应都在同一方向，从而造成弗吉尼亚有更多的人报告自己购买过武器。诸如这样的多重效应，也可能是反方向的，它会使得我们的判断变得更加复杂。最后，如果弗吉尼亚的枪支购买法不那么严厉，而马里兰的则可能严厉一些[5]，那么有可能的确有更多的弗吉尼亚的居民实际上已经购买了枪支，而在马里兰，可能购买枪支的人比报告的数字更多，但因为是非法的，所以不愿意报告。在这一案例中，不同的调查结果可能一部分是由两州人口之间存在的真实的差异所引起，而另一部分则是由两州居民对报告这一行为所持的态度不同所引起。诚然，试图在不同调查之间的结果进行比较，可能会引起很多的麻烦，并使问题变得更复杂，但是如果不这样做，对存在的差异视而不见，就有可能导致调查产生的结论得不到真正的支持。

在所有的例子中，读者或数据分析者都必须对调查的方法有详细的了解，以能恰当地使用调查的结果，这一点是十分清楚的。因此，一个认真的研究者有义务提供那些对调查的质量或对调查结果的解释会有影响的有关调查方法方面的重要信息。有些这样的信息如抽样误差，假如可能的话，最好能提供量化的信息，而其他有关方法论的信息，一般是描述性的就可以了。如果我们注意到各种方法方面的决策及其结果的清单篇幅很长，要对它们全部加以阐述，不仅费用很高，而且时间很长。我们可以退而求其次，择其要点报告，既不要包罗万象，什么都报告，也不要简单一提，几乎什么都不报告。

方法论报告的主要内容

表 10.2 列出了许多经常包括在一个方法论报告，有时也称之为误差概览或质量概览中有关调查的各种特征。我们先用一些简明的例子分题项，一一说明方法论报告中可能应该包括的内容，然后指出在某一特定的调查中，在这些方面进行决策时应该注意的事项，说明哪些问题是写报告时可取的或基本的，哪些问题是不太可取的或不太重要的。

样本设计

方法论报告经常都从抽样设计的讨论开始。这一节我们将为报告其余部分的讨论提供一个框架，同时也为后面有关抽样框、加权、估计和其他问题的讨论提供一个背景。此外，我们也希望告诉读者，设计对调查目标的最终实现至关重要。这一节至少应该包括调查目标总体的操作定义、调查主要目的的一般描述和采用的抽样方法的主要性质，例如是否采用了分层、聚类或多阶选取方法等。

表 10.2 质量概览

抽 样

- 抽样设计
- 抽样误差
- 抽样框存在的问题:覆盖面

数据收集

- 工具的设计和改进:测试特殊的方法、认知方法
- 数据收集:调查员培训、实际调查回拨质量控制/监控(校验)
- 回答偏倚:问卷问题
- 单位无回答:回答、拒绝回答和合作率;回答人筛选
- 特殊的程序:未回答问题的被调查人样本、拒访者转化、物质奖励
- 无回答偏倚:组间的无回答差别

估 计

- 加权和估计
- 题项无回答:编校和分配
- 数据录入和编码:录入校验和录入误差、开放式问题的编码

如果采用了分层抽样,那么分层是按比例的、不按比例的,或者是这两种分层方法的某种组合?如果采用的是不按比例的分层抽样,那么研究者应该注意这样做的主要原因是否是为了保证单独的子群体能有足够的个案数,以便进行子群体之间的比较或进行其他的分析。在层内,单元的选取是否等概?当然,我们也必须提供层的定义。

抽样误差

抽样误差是最为普遍的报告调查精度的量度。经常有人在报告中不恰当地用它来替代非抽样误差。这种做法固然是不可取的,但是这并不等于说我们的方法论报告可以不去谈抽样误差的问题,相反每一篇方法论报告,不仅应该都要对抽样误差加以阐述,而且如果可能,还应对所有的变量或关键的变量的抽样误差加以阐述。我们也可以把它们概括成一张表格,以便可将它同时用于许多个调查变量[6]。如果我们调查的样本是一个简单随机样本(或系统随机样本),那么这一工作是非常简单的。统计软件可以为我们提供简单随机抽样的抽样误差[经常被标为标准误差(standard errors)[7]]。通常它都作为一种例行的输出项,随其他计算结果一起输出。遗憾的是,在许多时候,样本设计与简单随机抽样相去甚远[8]。诚如我们所知,抽样设计经常涉及分层、聚类或多阶选取。与此同时,我们经常还需要考虑无回答和后分层的加权问题。

在遇有这样一些情况时，简单随机抽样误差将会过高地（有时可能会相当大地）估计测量的精度。

有什么样的备择方法可资我们利用呢？对许多设计而言，抽样误差可以通过教科书提供的数学公式进行计算[9]。然而，这些计算并非都是很简单的，即使对于中等大小的数据集，情况也同样如此。我们必须对这些公式进行编程，并把它们与测量的数据链接起来。

更为实际可行的方法是使用专门设计用于抽样误差（或设计效应[10]）的计算软件。这些软件已经考虑到了各种特殊的抽样设计和加权方法。在这方面使用最为普遍的软件（用于 IBM 的 PC 机）是 SUDAAN（SUrvey DAta ANalysis，意为调查数据分析）和 WesVar（http://www.westat.com/wesvar/）。两种软件的设计者分别为研究三角研究所和惠斯塔公司。这些软件的操作和使用的详细讨论，超出了本书讲授的范围，但是有兴趣的读者可以设法得到它们的操作手册和参加它们的短期课程。

即使这一费用（或编程的专门技术的水平）超出了我们手头握有的资源，我们还是可以退而求其次，关注查阅一下以往与我们的研究类似的研究设计的设计效应，并以此为依据，对自己的研究设计的设计效应做一些推断，甚至猜测。这种方法固然有相当大的风险，但是较之对方法论不作任何报告，或报告错误的简单随机抽样的抽样误差，这样的方法还是比较可取的[11]。

抽样框问题：总体差额（undercoverage）

就抽样框问题而言，方法论报告主要应该讨论的问题是总体差额问题，因为一般讲，任何其他的抽样框问题，如总体超额（overcoverage）和多次性（multiplicity）问题可能需要在抽样、数据收集或估计阶段做处理。我们可以做的诸如这样的陈述可能相当一般，但是仍然可能含有相当多的信息，例如“电话簿已经出版 8 个月，因此可能遗漏某些新住户，但具体比例不太清楚”。如果在某一组特定的变量上，我们注意到长期居住的住户和新来的住户之间存在着某些差异，那么诸如这样的一种陈述非常有用。即使我们的调查结果的读者在读了这样的陈述之后，并不能确定估计值中偏倚的大小，但至少可以对偏倚的方向有所了解。同样，如果我们必须在学年中使用一份学生的名册作为抽样框，那么我们也许会做这样一个注释：“该名册没有包含春季注册的学生，根据从大学管理部门得到的去年的有关数据，春季注册入学的学生的比例约为 10%。”报告中包含诸如这样的陈述的重要性取决于两个方面：总体差额的百分比和它与我们研究的变量之间可能存在的关系。当然，这样的陈述只不过是一些比较粗糙的估计。如果我们可以知道总体差额偏倚的方向和数量就更为理想。但是如果我们只知道其中的某些方面，总比什么都不知道要好。

在编写方法论报告的时候，我们不应当忽略我们的调查中做得比较好的

地方。例如我们可能设计编制了一个用于随机数码拨号(RDD)的、相当完整的、不存在抽样框总体差额问题的抽样框。无疑这在设计上是一个很大的优点,因而我们应该在方法论报告中的适当地方对它加以阐述。

工具的设计和改进:试调查和一些特殊的方法

方法论报告应当对试调查的次数和类型,以及是否仅限于常规性的,还是也包括了认知测试、行为编码、焦点组访谈、专家组访谈或这些方法的组合等加以说明。

即使经过了试调查,如果我们怀疑某些题项仍然存在某些问题,这样的情况我们也应该在报告中进行讨论。我们来回忆一下学生调查中的以下问题:

在大多数喝酒的时候,您平均一次喝多少?

而情况相当可能是,通过认知试调查,我们得知被调查人喝酒的情形并不是很有规律的,例如可能只是在遇有重大的社会事件时才痛饮一番,因而很难计算平均饮用量。问题在于,即使我们已经认识到了这样一个问题,我们也不能把这一问题取消,因为这个问题对我们的分析十分关键,因此我们别无选择,必须保留。而使使用我们调查结果的读者了解那些偶尔饮酒的人的测量误差可能比那些每晚都喝一次酒,或在星期五下午的"快乐时光(happy hour)"喝一次酒的被调查人的可能要大一些是很有用的。而对于上面提到的这样或那样的测量问题,有时研究者不愿过多提及。但是苏德曼(Sudman,1976)认为,样本设计偏倚问题的讨论,不但不会贬低调查的价值,相反,偏倚类型的详细讨论还会大大提高调查数据的使用价值,因为它使人们知道了调查结果的局限性是什么。此外,我们的报告究竟应该包含多少诸如这样的信息,则应取决于我们对诸如这样的一些不足对关键变量的可能的影响或读者希望对我们的调查优缺点和结论的普遍性的了解程度的判断。

数据收集:调查员培训、数据收集方法和质量控制

方法论报告应该说明数据收集的时间(如 2004 年 2 月到 4 月)。这样的说明似乎比较琐碎,但如果在某些变量存在季节效应(如在调查问及娱乐活动时),或在某一公共事件发生可能对调查结果有所影响(如在我们的犯罪调查的数据收集期间发生了一个广为人知的可怕的犯罪事件,有可能降低人们选择不监禁判决这一类答案的愿望)时却很重要。

报告也应对调查员的培训情况做一个简短的总结。与此同时,也应对在培训时,我们给调查员指出的那些用以处理那些特别的难题或问题的说明做一个简短的介绍。这样的介绍必须逐题进行,不可笼统。例如,我们来考虑一下下面的问题:

假如一个年龄为____的人被证明贩卖价值____的可卡因,且犯前科____

次，那么您认为他应该被判入狱，要求他每月向假释官报告一次、每天向假释官报告一次，还是需要受到监控，以便我们掌握他的所有行踪？

1. 投入监狱　　2. 要求每月报告一次

3. 要求每天向假释官报告　　4. 进行监控，掌握他的全部行踪

8. 不知道

对于这样一个题目，被调查人可能会提出各种各样的问题。如我们的确有这样的监控设备吗？假如他没有向假释官报告情况会怎么样？等等。如果前面培训的时候，我们已经给了调查员有关的指示，告诉他们在遇到前一个问题时应该说"是的，确实有这样的装置"，而在遇到后一个问题时应该说"假释官将根据具体发生的情况决定怎么做"。这些情况都应该在报告中有所提及。如果诸如这样的特殊说明有许多，那么比较可行的办法是将所有这样的说明全部收入一个附录，附在报告的末尾。

报告也应该提及回拨的次数，以及回拨日程安排的一般规则，同时拨号监听或效验的百分比也应该有所交代。许多调查计划虽然对 10% ~ 15% 的调查进行了监听，但在我们的印象当中，这一数字实际上似乎也包括了许多只进行了部分监听的调查。

回答偏倚：问卷问题

如果我们有理由（例如我们在调查监听过程中发现的情况，或从调查员的评论中听到的情况）认为，某些特定的问题会给被调查人的回答带来问题，这样的情况也应该在报告中提一下。例如，可能有许多被调查人对我们为某一问题提供的各类答案感到费解或反感。在大多数情况下，诸如这样的问题都会在问卷的设计阶段得到确认和解决，但是，如果在我们尽力做了各种改进之后，问题似乎仍然存在，那么作为研究者，我们有义务对这样的情况进行报告。

单位无回答：回答率、拒绝率和合作率

与调查质量的其他指标不同，调查的各种率从来都不是可有可无的报告内容。这些内容不是要不要报告的问题，而是怎么样报告的问题。在各种调查率中，其中回答率、拒绝率、合作率和完成率是为不同的研究者所普遍采用的，但使用的意义却不尽相同。尽管对这些率的计算和表达方式不乏为人推崇的版本（参见 Council of American Survey Research Organizations, 1982; Groves, 1989; Groves and Couper, 1998; Hidiroglou, Drew, & Gray, 1993），但是它们并未为人们所普遍接受。有关这方面的最近的尝试刊载于美国舆论调查研究会的出版物中（American Association for Public Opinion Research, AAPOR, 2004），它推荐使用用于电话调查和面对面调查的有关这几个调查率

的定义和计算公式。我们建议使用一组较为简单明了的调查率,但是更为重要的是,我们敦促研究者都应该确切的交代自己的调查中各种调查率的具体定义和计算方法、公式中样本部署码的取舍规则。

在一般人口总体的电话调查中,主要的拨叫结果将被归为三类:非住户(或其他的不合格户)、住户和住户状况未知。非住户的数目十分清楚,它包括未在运行状态的电话、营业性或其他组织机构的电话、旅馆、集体住户等[12]。住户这一类则包括了已完成的调查、拒绝调查和各种无法联系的类型。而所谓无法联系是指已经确定的(非拒绝调查的)住户中选取的被调查人,在进行了规定次数的试拨之后,因各种原因,如外出旅行、度假、伤病或应答电话始终都是应答机等始终无法取得联系的住户。此外,有些无回答是因为住户没有讲英语的成员所致。如果根据调查的定义,这些个案都是属于合格的个案,那么它们都必须包括在各种调查率的计算中。最后,即使在进行了相当多次的回拨之后,还是会有一些(相对较少)电话号码的住户的实际状况无法确定。无论把这些电话号码都计为合格的还是都计为不合格的都是不恰当的。恰当的做法是采取一种比较合理的折中的计算方法,把它们中间一定比例的号码计为合格的住户。而我们已有的有关这样的计算的最好指南是我们能够确定这样的号码中住户号码的百分比(参见 AAPOR, 2004)。

现在我们再回到表 9.3。数字显示,在 1 816 个初始样本号码中,有 702 个是各种类型的非住户,1 061 个是住户,53 个在拨叫了 20 次之后仍然无法确定号码的性质。我们建议,按惯例,如表 9.3 所示,使用 1 061 个已经确认的住户作为计算回答率、拒访率和没有取得联系率的下限。

为了计算上限,我们注意到 1 763 (1 061 + 702)个电话号码可分为住户和非住户两类,其中大约 60% (1 061/1 763)为住户。如果采用这同样的百分数,那么我们将要在分母中加上 32 个个案,这样回答率便从 78% (824/1,061)变为了 75.4% [824/ (1 061 + 32)]。这一方法给出了有关以电话号码住户状态未知的假定为依据的回答率的范围。

一般情况下,最后阐述的都是合作率。这种比率的计算是以确认的合格的调查数除以调查数与拒访数、部分调查数和其他的户数之和。这一比率的值表示我们的调查在那些实际已经联系上了,且户内有合格的被调查人的户中完成的好坏。在两种情况下这一比值特别有价值。第一,在数据收集期间,我们希望对实际正在进行的调查的进展情况进行跟踪。在调查进行期间,出现了数量相当多的无法联系的现象,因而回答率必定很低,这时成功率便不再是一个很好的测量事情调查实际进展情况的指标。其次,有时数据收集的时间是如此短促,以致我们可以实现的回呼的次数十分有限。在调查试图测量公众对某一新闻事件的反应时,这样的情况就有可能发生。由此而导致的结果又是数目众多的无联系。在这两种情况下,合作率将是一个比回答率更好

的测量调查是否成功的指标。

特殊的手段：无回答的被调查人样本和拒访转化

任何作为数据收集工作的一部分的特殊手段和方法，如拒访转化或对未回答问题的被调查人样本进行的特殊补充联系都应该予以阐述。我们应该对使用的方法及其最终达到的（或未达到）效果加以说明。此外，如果有相当数量的个案（比如说 15% 或更多）是通过这样的工作才得到的，那么简要地说明一下这些个案与样本中的其他被调查人，在那些关键变量上（在统计显著性上[13]）有什么不同是非常有用的。

无回答偏倚：无回答模式和权的使用

如果单位无回答似乎是非随机地分布在整个总体中，尤其是在分布的模式与那些实质性的变量相关的时候，那么报告应对这些情况加以讨论。例如，在一个全州范围的调查中，如果城市被调查人的回答情况大大差于农村，那么我们就应该在报告中说明这一情况。即使对这样的情况用加权进行了修正，我们也必须进行说明，也许这时，它比没有加权更加应该进行说明。如果我们用加权来修正无回答、不等概选择或后分层，我们应该对此予以说明，尽管我们不必对加权本身在方法论报告中展开详细的讨论。

估　计

在大多数小规模的调查中，估计方法——即将样本估计值转变为对应的总体参数估计值——都是非常简单的，甚至是不言自明的。在情况并非如此时，我们应将估计采用的数学公式列在方法论报告中。

题项无回答和分配

在大多数报告中，都未曾对题项无回答问题做任何阐述，即使阐述的话，也无非是三言两语而已。方法报告必须包括题项无回答问题。报告的重点应该是是否存在异常的情况。例如，譬如说，任何题项只要它的无回答超过了 5%，都应该在报告中提及。但是，与单位无回答一样，对于我们的研究发现而言，最重要的问题是题项无回答的分布。诚如我们所知，题项的丢失常常是非随机的，因而可能会对我们的分析有一定的影响。

分配（imputation）是用构建的数值替代那些无回答或答案与同一调查的其他答案不一致的题项的原来值。在大多数情况下，构建值是以数据集中的其他个案的数值或同一个案的其他变量的值为根据的。在小规模调查中这样的做法并不少见，但是我们不主张初出茅庐的研究者这样做。如果题项无回答问题的严重程度已经危及我们要做的关键性分析（譬如对某一特定的子群

体)，那么分配法是值得一试的。在这样的情况下，读者应该参阅卡尔登(Kalton,1981)，或格罗夫斯、迪尔曼、埃尔丁格和利特尔(Groves, Dillman, Eltinge, & Little,2002)等人有关备择分配法的讨论。而我们则认为首选的方法是改进题项的测试方法，最大限度地减少无回答问题，防患于未然，而不是在事后才对无回答问题进行各种纠正。

数据录入和编码:录入误差、开放式问题

数据录入的误差率或估计的误击数在总击键数所占的百分比一般是不报告的，除非存在某种较为普遍的问题。在为某些题项编码时遇到的特殊困难应该进行报告。如果这样的情况确有发生，一般都发生在开放式题项上。在许多情况下，如果研究者已经对试调查做了严密的监控，并对试调查的个案做了实际的编码，那么他们可能就不必在方法论报告中对诸如这样的问题多费口舌。

费用和意外事件:计划外开支预算

我们已经对完成一个调查所必需的一般设计问题和详细的操作方法有了相当的了解，且对这一期间可能会发生什么样的意想不到的事件也已有所了解，现在我们转而对调查费用的问题做一些更为详细的讨论。在这里，我们是在广义的意义上使用费用一词的，它泛指资料或服务设施(如为电话调查的调查员和数据录入所需的设备)和时间投入(如无酬研究者)等，即使它们不需要花钱。我们仍以马里兰犯罪问题调查为例。

预算不可能是完全科学的(众所周知，政府的开支总是超过预算)，但却是不可或缺的。在执行一个调查的过程中，如果对所需的主要费用不甚了了，或大大低估，其后果可能是灾难性的。预算严重不足将导致我们在达到预期的样本数量之前，或在达到为得到足够好的回答率所需补充调查数之前，被迫结束数据收集工作。

为了制订好调查的预算，一个简单易行的方法是列出为完成一个调查所需的所有步骤。进而弄清每一步骤所需的资源及它们是否需要一定的开支。即使一个不需要付钱的工作，我们也必须把它列出，它不仅可能会对我们的实际计划很有帮助，而且也会对我们如何处理某些其他的工作有一定影响。例如，研究者本打算亲自打印所有的问卷草稿，从而可以节省试调查中这一部分工作的费用。然而在列出了这一工作及其他“免费”的工作所需的时间之后，研究者发现，还是把自己的时间用在其他方面更好，于是改而将问卷草稿的打印等工作转交给了秘书或打印社。

如果我们知道各种工作的费用，那么使用类似表 10.3 所列的那样的任

务/费用图可以使我们很快就能制订好一个预算。发现某些项目的确切费用也许是很困难的。为了做到这一点,我们可能需要做一些估计(例如电话费)。但是以下这些程序将迫使我们了解我们的预算的不确定性在什么地方,在什么地方我们必须考虑到会有意想不到的开支,给自己留有必要的余地。它也将迫使我们估计调查真正需要的时间和费用。初涉调查领域的研究者最常犯的错误是低估做上乘的调查所需要时间和精力。

表 10.3　调查工作和费用预测:马里兰犯罪调查 *

任务	项目	费用
数据收集前		
编写调查目的提纲	8 小时,研究者起草	免费
编写问卷主题提纲	24 小时,研究者起草	免费
	16 小时,同事评议	免费
查阅问题实例	24 小时,学生助理	$ 258
	4 小时,研究者评议	免费
编写新问题	16 小时,研究者起草	免费
	16 小时,学生助理评议	$ 192
打印问卷草稿	4 小时,秘书	$ 38
获取抽样框	8 小时,学生助理	$ 96
	2 小时,秘书	$ 19
	购买计算机化的抽样框	$ 650
选取试调查样本	2 小时,研究者	免费
	6 小时,学生助理	$ 72
准备培训材料	16 小时,研究者	免费
	20 小时,专业培训员	$ 360
	10 小时,秘书	$ 95
	培训材料	$ 50
培训试调查调查员	16 小时,专业培训员	$ 288
	160 小时,(10)调查员	$ 1 280
	48 小时,(3)督导员	$ 480
	2 小时,研究者	免费
试调查和任务报告书	4 小时,专业培训员	$ 72
	40 小时,(10)调查员	$ 320
	12 小时,(3)督导员	$ 120
	4 小时,研究者	免费
检查结果	6 小时,研究者	免费
	16 小时,学生助理	$ 192
修正和复印问卷	16 小时,研究者(复印)	免费
	24 小时,学生助理	$ 288
选取主样本	8 小时,学生助理	$ 96

续表

任务	项目	费用
数据收集		
主调查调查员培训	16 小时,专业培训员	$ 288
	160 小时,(10)调查员	$ 1 280
	48 小时,(3)督导员	$ 480
	2 小时,研究者	免费
800 个 15 分钟的电话调查和监控(5 星期的数据收集)	700 小时,(10)调查员	$ 5 600
	200 小时,(3)督导员	$ 2 000
	电话费	$ 1 620
数据检查	25 小时,学生助理	$ 300
做拒访者的工作	20 小时,学生助理	$ 240
	10 小时,督导员	$ 100
数据收集后		
将数据录入计算机文件	200 小时,录入员	$ 1 300
	20 小时,督导员	$ 200
数据分析	20 小时,研究者	免费
	40 小时,学生助理	$ 480
写报告	20 小时,研究者	免费
	40 小时,学生助理	$ 480
	30 小时,秘书	$ 285
人工小时总计	2 155	
费用总计		$ 19 529

* 估计值以一个调查时间为 15 分钟的 RDD 电话调查为依据。

我们从分三个阶段列出调查所需的各种工作开始来制订我们的工作任务表:必须在数据收集之前完成的工作任务、数据收集期间应该完成的工作任务和数据收集工作完成之后应该完成的工作任务。在表完全做好之前,有些工作可能要反复几次,因为在有些条目的制订和费用的填写过程中,资源的具体分配可能会有一定变动,这一点与我们前面引用的问卷草稿打印的例子颇为相似。

纯粹就必须完成的工作任务的数目而言,为完成一项调查所需事先完成的各项准备工作可谓数目繁多。尽管最大量的工作时间将花费在电话调查本身上,但许多工作任务必须在调查正式开始之前完成。我们也应该了解,电话调查的绝大部分费用都花在了人工身上,只有一小部分用于材料和其他开支。虽然在人工花费较少的邮寄式调查中,这一费用平衡模式可能会有所变化,但是前期的工作仍然需要相当多的人工。因而这种类型的调查人工费仍然不菲。而从这一角度看,我们可以了解,尽管邮寄式调查的其他开支低于电话调

查，但是设计和组织实施的费用仍然是相当高的。

虽然表 10.3 列出的数字是例证性的，但是它们基本上适用于大学的调查组织进行的调查。当然，如果“免费”的时间更多——例如由学生来做所有的调查——那么费用（并非人工）会更低一些。大学的调查常常作为课堂作业进行，调查工作同时由调查组织的雇员和课堂里的学生担任。这时，不用将全部课时用于实际调查，样本便可达到合理的数量。作为一个练习，大家不妨考虑一下，在这种“免费”和付费的混合型调查中，费用应该如何计算。

大多数任务的费用和时间是可以事先估计，或在请教某些曾经做过调查的人之后能大致明确的（在做调查计划时，这种做法总是一种明智之举）。但是调查本身所需要的时间是一个例外。一种用于估计这一时间的、使用颇为普遍的方法是做如下所示的计算：

$$\text{总调查时间} = \frac{\text{每个调查所需的时间} \times \text{样本量} \times \text{I. E. R.}}{60}$$

公式中唯一需要说明的部分是调查员效率（interviewer efficiency rate，I. E. R.）。效率是（以过去的经验为依据）每分钟的实际调查工作所花费在辅助性工作（拨叫无结果的号码、处理样本、休息等）上的时间估计数。在本例中，为完成 800 个 15 分钟的调查，共用 700 个调查小时，I. E. R. 为 13.5。这就是说，每一分钟实际调查大约需要为其他事务（如与被调查人联系、再次拨叫拒访者，等等）耗费 2.5 分钟。因为 I. E. R. 对恰当地计算调查时间至关重要，因此我们希望大家能在这方面多多请教资深的调查研究者。

我们之所以不厌其详地反复阐述那些旨在降低资源误差的各种调查程序方法，其目的在于使读者明了，为了较好地完成调查任务，虽然我们必须涉及众多的细节，但其中大多数在概念和操作上都是很简单的。问题的关键在于认真地计划好资源和时间的实际配置，并区分各种任务的轻重缓急。

进一步阅读的推荐

限于篇幅，本书涉及的内容主要集中在那些进行小规模的调查时所必须做出的决定和方法步骤，它对初涉调查研究的新手比较适合。虽然我们希望诸如这样涉及面比较窄，但比较集中的讨论，能更有助于那些调查方面的新手解决他们所遇到的许多实际问题，但是我们也认识到，这样一种叙述方式不免有把调查研究领域中的专著扭曲为简单的“菜谱”之嫌，进而对学科本身和怀有浓厚的求知欲的读者造成一定危害。

近年来，在调查研究领域，研究者和实际工作者携手努力，开拓创新，经典佳作屡屡面世，见证了调查研究领域的各个方面在方法论研究和概念创新方

面的与时俱进。在问卷设计方面的研究和应用的佳作有佩恩的《提问的艺术》(*The Art of Asking Questions*, Payne, 1951)、苏曼和普雷舍的《态度调查的问题和答案》(*Questions & Answers in Attitude Surveys*, Schuman & Presser, 1981),及苏德曼和布拉德伯恩的《提问》(*Asking Questions*, Sudman & Bradburn, 1982)。这些著作在认知心理学调查研究的工具设计和数据收集方法中得到了广泛的应用。自佳宾、斯塔拉夫、塔纳和塔伦乔的《调查方法论的认知问题》(*Cognitive Aspects of Survey Methodology*, Jabine, Straf, Tanur, and Tourangeau, 1984)问世以后,再经塔纳的《问题的问题》(*Questions About Questions*, Tanur, 1992)和由施瓦兹、苏德曼和他们的同事组织的研讨会系列论文集[14],这一领域在过去的四年间又有了长足的进步。

许多早期的有关抽样理论和方法的经典著作,如凯思的《调查抽样》(*Survey Sampling*, Kish, 1965)和考克朗的《抽样技术》(*Sampling Techniques*, Cochran, 1977),连同苏德曼的实用性指南《实用抽样技术》(*Applied Sampling*, Sudman, 1976)一直在有关专著,如沙恩德尔、斯文森和芮特曼的《模型辅助的调查抽样》(*Model-Assisted Survey Sampling*, Sarndal, Swensson, & Wretman, 1992)中被广为引用。即使像本书开列的这样一张简短的清单,如果遗漏了数据收集方面的重要著作,如迪尔曼的《邮寄和电话调查:总体设计方法》(*Mail and Telephone Surveys: The Total Design Method*, Dillman, 1978);测量误差方面的重要著作,特纳和马丁的2卷本《调查主观现象》(*Surveying Subjective Phenomena*, Turner & Martin, 1984);一般设计问题中的重要著作,凯思的《研究的统计设计》(*Statistical Design for Research*, Kish, 1987)和涵盖了调查研究的多个领域的重要著作,格罗夫斯的《调查误差和调查费用》(*Survey Errors and Survey Costs*, Grove, 1989),也不免有玩忽职守之嫌。最后,我们认为有必要提及的是,由美国主要调查研究专业协会组织的研讨会,在有关调查研究的特定领域出版了几本很有价值的专论。这些专论包括格罗夫斯等人的《电话调查方法论》(*Telephone Survey Methodology*, Groves et al., 1988);贝摩、格罗夫斯、卡尔登、来博格、马休斯和苏德曼等人的《调查的量度误差》(*Measurement Errors in Surveys*, Biemer, Groves, Lyberg, Mathiowetz, & Sudman, 1991);卡斯普斯基、登肯、卡尔登和辛等人的《期群调查》(*Panel Surveys*, Kasprzyk, Duncan, Kalton, & Singh, 1989)和来博格等的《调查的量度和处理的质量》(*Survey Measurement and Process Quality*, Lyberg et al., 1997)。当然,除此之外,还有许多杂志刊登的论文和会议的论文集,数量众多,不一而足。

自本书第1版问世将近10年以来,又有许多新的文献面世,其数量之多,使我们还来不及开始对它们做全面的概括。而本书第1版所提供的资源和信息不仅仍然十分重要,而且至今仍可沿用,因而本书的第2版推荐的新的信息

资源仅限于调查研究方面出现的新领域和新问题。其中有关互联网调查的有迪尔曼的《邮寄和互联网调查：设计方法精选》(*Mail and Internet Surveys: The Tailored Design Method*, Dillman, 2000)和库珀等人的《计算机辅助的调查信息收集法》(*Computer Assisted Survey Information Collection*, Couper et al., 1998)。理解调查答案问题的认知心理学新著则有塔伦乔、利普斯和拉辛斯基的《调查答案的心理学》(*The Psychology of Survey Response*, Tourageau, Rips & Rasinski, 2000)、舍金等人的《认知和调查研究》(*Cognition and Survey Research*, Sirken et al., 1999)和普雷舍等人编辑的《调查问卷的测试和评估方法》(*Methods for Testing and Evaluating Survey Questionnaires*, Presser et al., eds., 2004)。在有关如何提高回答率和调查质量方面的专著则有格罗夫斯等人的《调查的无回答问题》(*Survey Nonresponse*, Groves et al., 2002)及贝摩和来博格的《调查质量导论》(*Introduction to Survey Quality*, Biemer & Lyberg, 2003)。

虽然这些内容广泛、数量众多的阅读材料的主要阅读对象是专业的调查研究工作者，但是我们还是要认真地把它们推荐给诸位。因为即使是入门不久的新手，这些资料都会对他们调查过程中的每一方面都有很好的参考价值。尽管它们都含有很多的理论，但是它们仍然有很高的可读性。此外，它们还会在给我们以很多实用的忠告的同时，帮助我们更好地理解调查研究中的许多基本概念。而这一切都是其他阅读材料所无法给予的。即使是那些只花了几天时间浏览了这些著作的学生，也会因为对学科的真谛有所洞察和对一个正在继续扩展自己的领地及不断引进新的思想和技术的研究领域的四射活力有所感受而受益匪浅。

注 释

1 有时用无回答的被调查人的样本来估计这一差别是可能的。在一般情况下,我们使用某种备择的数据收集方法对诸如这样的样本进行调查。虽然这样做会导致更多的开支,但是却可使回答率达到更高。

2 作为一个练习,读者应该用这些百分比计算权数,并重制表 10.1 的第二部分。

3 虽然标准的统计软件如 SPSS 和 SAS 都可以计算这些加权的估计值。不过为了更好地理解加权对结果的影响,我们仍然应该自己动手进行几种计算。

4 作为一个练习,读者应该自己动手计算置信水平为 68%、90% 和 99% 的标准误差。注意这些计算既不用考虑有限总体的修正问题,也不必考虑设计效应问题,只要考虑这两个因子对抽样误差的估计值的可能影响。

5 此外,我们假定枪支是人们买作自卫的主要武器。

6 诸如这样的一张表格系用众多的研究变量的抽样误差的平均值构建的,而这些均值在表中分不同的百分比划分,90/10、80/20、70/30、60/40 和 50/50 列出。这样我们可以找到某些结果表示为二分的百分比的特定问题的抽样误差的大致的值。作为一个练习,请读者自行计算子样本量为 100、200、300 和 400 的、95% 置信水平的抽样误差(假定抽样方法为简单随机抽样)。参见表 7.1。

7 标准误差是标准差除以样本量的平方根。一般讲,方法论报告应该给出对应于 90%、95% 或 99% 置信水平的标准误差。

8 记住,等概选择法(equal probability of selection method)得到的 epsem 样本并不总是等价于简单随机样本。许多非常复杂的抽样虽然最终都会导致选择的等概,但是这样的抽样设计的抽样误差与简单随机样本的有很大不同。

9 例如,可参见凯思的有关著作(Kish , 1965)。

10 设计效应是复杂的抽样设计的方差与同样本量的简单随机样本方差的比率(Kish, 1965)。一旦有可资使用的设计效应的值,我们便可将它们转换为简单随机抽样的标准误差,以对复杂抽样设计进行修正(参见第 7 章)。然后,这些转换的抽样误差便可以上面讨论的方式使用,生成适用于特定的样本设计的抽样误差表。

11 如果方法简介没有讨论这一问题,那么有些比较单纯的读者有时就会误以为使用的是简单随机样本的抽样误差。

12 居住单位类型的完整定义请参见美国普查局 1990 年人口和住房普查:说明书第一部分的有关内容(Washington, DC: U. S. Government Printing Office, 1992),或登录http://fact-finder. census. gov。

13 这样的比较使用交互表的卡方检验便足够了。

14 其中有希博勒、施瓦兹和苏德曼(Hippler, Schwarz, & Sudman , 1987)和苏德曼(Sudman, 1992, 1994)。

附录A

马里兰大学本科生调查

Undergraduate Student Survey

本调查为调查人保密，您不必在问卷上签名。

除了特别说明的问题之外，请在最能反映您的观点的那一个答案前的号码上画圈。

图书中心

1. 本学期，从一月份以来你是否常到图书中心买书？

   1. 只有一两次　　2. 每星期不到一次　　3. 每星期一次
   4. 每星期一次以上　　5. 从来没有去过⇒第 10 题

2. 这学期你是否在图书中心购买了新的教科书？

   0. 没有⇒第 4 题　　1. 购买了

3. 你对新教材的质量、价格和选择性是否满意？

	很满意	比较满意	不太满意	很不满意
a. 质量	1	2	3	4
b. 价格	1	2	3	4
c. 选择性	1	2	3	4

4. 这学期你是否在图书中心购买过一般的文具用品，如钢笔和纸张等？

   0. 没有⇒第 6 题　　1. 购买了

5. 你对一般文具用品的质量、价格和选择性是否满意？

	很满意	比较满意	不太满意	很不满意
a. 质量	1	2	3	4
b. 价格	1	2	3	4
c. 选择性	1	2	3	4

6. 这学期你是否为你上的课，通过图书中心特别预定了有关书籍或其他材料？

   0. 没有⇒第 9 题　　1. 是的

7. 你对这些特别预定服务是否满意？

   1. 很满意　　3. 不太满意
   2. 比较满意　　4. 很不满意

7a. 满意的主要原因
   1. 时间
   2. 选择性

7b. 不满意的主要原因
   1. 时间
   2. 选择性

3. 价格　　　　　　　　　　　　3. 价格
4. 服务　　　　　　　　　　　　4. 服务
5. 其他,请具体说明______　　　5. 其他,请具体说明______

8. 你认为哪一件事情对图书中心特别预定服务工作的改进最重要?

____

____

9. 一般讲,你对图书中心的工作人员给顾客的帮助、礼貌及对产品和服务的了解怎么评价?

	好	非常好	一般	差	没有经验/没有看法
a. 给顾客的帮助	1	2	3	4	8
b. 礼貌	1	2	3	4	8
c. 对产品和服务的了解	1	2	3	4	8

教　学

10. 一般讲,你进入马里兰大学以来,你与你的指导学术导师会面的频率是多少?
1. 从没有见过　　　　　　2. 一学期不到一次
3. 一学期一次　　　　　　4. 一学期两次
5. 一学期三次　　　　　　6. 一学期三次以上。

11. 你现在的指导教师是?
1. 专职的指导教师(不在编的全职指导教师)
2. 同辈(本科生)　　　　　3. 研究生助理
4. 教员　　　　　　　　　5. 秘书
8. 不知道　　　　　　　　9. 目前没有指导教师

12. 你喜欢什么样的指导教师?
1. 专职的指导教师(不在编的全职指导教师)
2. 同辈(本科生)　　　　　3. 研究生助理
4. 教员　　　　　　　　　5. 秘书
6. 无所谓

13. 你是否有过由同辈担任的指导教师?
是　　　　　　　　　　　否⇒跳答第 14 题

13a. 你是否认为同辈指导教师至少应该与自己同一年级?
1. 应该至少同一年级　　　2. 什么年级无所谓

14. 在下列各项工作中,你认为你的指导教师对你的重要程度是?

	很重要	比较重要	不太重要	根本不重要
1. 有关毕业必备的条件的信息	1	2	3	4
2. 推荐介绍校园内的其他服务设施	1	2	3	4
3. 学业指导	1	2	3	4
4. 决定每学期选修的课程	1	2	3	4

5. 在计划本科生项目时给予自己的咨询 1 2 3 4

15. 一般讲，自从你进入马里兰大学以来，你得到的指导对你是否有帮助？

1. 很有帮助　　2. 有一定帮助

3. 不太有帮助　　4. 根本没有帮助

16. 那些帮助你选好选修课的指导对你是否有帮助？

1. 很有帮助　　2. 有一定帮助

3. 不太有帮助　　4. 根本没有帮助

5. 没有经验

17. 那些帮助确定你学业目标的指导对你是否有帮助？

1. 很有帮助　　2. 有一定帮助

3. 不太有帮助　　4. 根本没有帮助

5. 没有经验

18. 你现在的学习成绩平均在？

1. 3.6 ~ 4.0　　2. 3.0 ~ 3.5

3. 2.5 ~ 2.9　　4. 2.4 或以下

19. 与其他同年级的学生比较，你认为你的平均成绩是？

1. 大大高于平均成绩　　2. 略高于平均成绩

3. 等于平均成绩　　4. 略低于平均成绩

5. 大大低于平均成绩

20. 与其他同年级，且同种族的学生比较，你认为你的平均成绩是？

1. 大大高于平均成绩　　2. 略高于平均成绩

3. 等于平均成绩　　4. 略低于平均成绩

5. 大大低于平均成绩

21. 以下问题是否在你身上发生过？（如果你不知道，或不能确定，请在 8 上打圈）

	经常	有时	很少	从来没有	不知道
1. 因为我的种族，得到过很不公正的分数	1	2	3	4	8
2. 与我同一种族的教师总是希望我比其他种族的学生强一些	1	2	3	4	8
3. 那些其他种族的教师总是希望我比他的种族的学生差一些	1	2	3	4	8

黑人学生问题

（这一部分的内容有关其他种族的人们是如何对待黑人学生的）

22. 据你的观察，以下反黑人的行为是否经常在校园中发生？

	经常	有时	很少	从来没有发生
1. 学生发表的种族言论	1	2	3	4

2. 因为种族而被排斥在社会事件之外	1	2	3	4
3. 教职员发表种族主义言论	1	2	3	4
4. 奖学金和荣誉上的种族歧视	1	2	3	4

23. 你毕业的高中的种族构成情况是？
   1. 黑人和白人各占大约一半
   2. 黑人为主
   3. 白人为主

酗酒和吸毒

24. 学校下列违规行为的最高处罚是？

	留校察看	毒品检查	取消住宿资格	停学	开除	不知道
1. 携带或吸食非法的毒品	1	2	3	4	5	8
2. 未成年携带酒精饮料	1	2	3	4	5	8
3. 出售或散布非法的毒品	1	2	3	4	5	8
4. 将酒精饮料提供给21岁以下的人	1	2	3	4	5	8
5. 在校园内饮酒	1	2	3	4	5	8

25. 总的来讲，你对以下校园政策是否熟悉？

	很熟悉	熟悉	不太熟悉	不熟悉
1. 携带或吸食非法的毒品	1	2	3	4
2. 出售或散布非法的毒品	1	2	3	4
3. 毒品检查	1	2	3	4
4. 未成年拥有酒精饮料	1	2	3	4
5. 将酒精饮料提供给21岁以下的人	1	2	3	4
6. 在校园内饮酒	1	2	3	4

26. 根据案情的具体情况，有些学生认为对于携带和/或使用非法的毒品的处罚，可以用为期两年的随机毒品测试来替代暂停学业。

你认为这样的政策是否公正？（在下列一个数码上画圈）

很公正	公正	不公正	很不公正	没有看法
1	2	3	4	8

27. 下列问题你的态度是？

	支持	强烈支持	反对	强烈反对	根据情况决定
1. 在校园内秘密调查	1	2	3	4	9
2. 强行开除	1	2	3	4	9
3. 警察在宿舍内任意巡视	1	2	3	4	9
4. 立法取缔在足球赛期间饮用酒精饮料	1	2	3	4	9

28. 假定你想得到下面的物品，那么你认为是否在校园内很容易得到？

	很容易	容易	困难	很困难	不可能
1. 大麻/印度麻药	1	2	3	4	5
2. 可卡因/快克可卡因丸	1	2	3	4	5
3. 迷幻药	1	2	3	4	5
4. 安非他明(快速)	1	2	3	4	5
5. 酒	1	2	3	4	5

29. 在过去的12个月中你是否经常使用下列毒品？

	从未使用	1～2次	3～10次	每月一次	每周一次	每天一次
1. 大麻/印度麻药	1	2	3	4	5	6
2. 可卡因/快克可卡因丸	1	2	3	4	5	6
3. 迷幻药	1	2	3	4	5	6
4. 安非他命(快速)	1	2	3	4	5	6
5. 酒						

30a. 在大多数情况下，你喝酒的时候平均每次喝多少个酒？____

[一个酒＝一杯啤酒、一杯混合型酒精饮料或一杯葡萄酒]

□ 从来也没有喝过酒

30b. 如果你从来都没有用过，或去年没有用过非法的毒品，那么你没有使用的原因是什么？

（请在某一个答案上画圈）

1. 这有悖于我的信仰
2. 这违反法律
3. 其他人不同意
4. 难以得到
5. 担心有害健康
6. 没有得到预期的效果

7. 使用毒品的有害体验　　　　8. 没有这样的愿望

9. 其他：________________

31. 如果你去年没有喝过酒，那么你没有喝酒的原因是什么？
(请在某一个答案上画圈)

1. 这有悖于我的信仰　　　　2. 这违反法律

3. 其他人不同意　　　　4. 难以得到

5. 担心有害健康　　　　6. 没有得到预期的效果

7. 使用毒品的有害体验　　　　8. 没有这样的愿望

9. 其他：________________

被调查人基本情况

D1. 你是？

1. 男　　　　2. 女

D2. 你是？

1. 白种，非西班牙裔　　　　2. 黑人或非洲裔美国人，非西班牙裔

3. 墨西哥裔美国人、波多黎各或其他西班牙裔　　　　4. 亚裔美国人

5. 本土美国人、美国印地安人　　　　6. 其他种族

D3. 你目前是

1. 一年级新生　　　　2. 大学二年级

3. 大学三年级　　　　4. 大学四年级

D4. 在过了最近一个生日之后，你的年龄是？ ______岁

D5. 你现在在哪个学院注册？ ________________

D6. 在刚进入马里兰大学之前，你在什么样的学校上学？

1. 高中　　　　2. 从两年制专科学院转过来

3. 从四年制专科学院转过来　　　　4. 在休学之后返校

谢谢你填完了这份问卷。

你的答案将完全保密。

请将问卷用附上的信封寄回。

附录B

马里兰犯罪调查

Maryland Crime Survey

您好，我从马里兰大学给你打电话。我的名字叫________。我们正在为州最高当局做一个有关街头暴力犯罪问题的调查。我们需要与您家中年龄在18岁或以上、接下来最快要过生日的那个人谈话。请问那个人是哪一位？

1. 一般讲，您认为您的**邻里地区**的犯罪问题是很严重、比较严重、不太严重，或一点也不严重？

   1. 很严重　　2. 比较严重
   3. 不太严重　　4. 一点也不严重
   8. 不知道

2. 一般讲，您认为**本州**的犯罪问题是很严重、比较严重、不太严重，或一点也不严重？

   1. 很严重　　2. 比较严重
   3. 不太严重　　4. 一点也不严重
   8. 不知道

3. 去年，您认为您的邻里地区的暴力犯罪问题是更多了、更少了，或几乎一样？

   1. 更多了　　2. 更少了
   3. 几乎一样　　8. 不知道

4. 去年，您认为**本州**的**暴力**犯罪问题是更多了、更少了，或几乎一样？

   1. 更多了　　2. 更少了
   3. 几乎一样　　8. 不知道

5. 去年，您认为您或您本人的财产是否遭遇过犯罪行为的侵害，或犯罪企图的侵害？

   0. 没有[跳答第10题]
   1. 是的　　8. 不知道

6. 您个人是否受到威胁或武力的伤害？

   0. 没有　　1. 是的
   8. 不知道

7. 去年，您认为您家中的任何人是否遭遇到过犯罪行为的侵害，或犯罪企图的侵害？

   0. 没有[跳答第9题]

1. 是的　　　　8. 不知道

8. 他们是否受到威胁或武力的伤害？

0. 没有　　　　1. 是的
8. 不知道

9. 明年，您觉得您或您的家人是否可能遇到抢劫或侵犯？您认为这种可能性将是？

1. 很可能　　　　2. 有可能
3. 不太可能　　　　4. 根本不可能
8. 不知道

10. 明年，您觉得您的家是否可能被人破门而入？您认为这种可能性将是？

1. 很可能　　　　2. 有可能
3. 不太可能　　　　4. 根本不可能
8. 不知道

下面我们想问的问题，有关您或您的家人在防范犯罪行为时采取的一些措施。

11. 去年内，您是否参加了社区防止犯罪项目？

0. 没有　　　　1. 是的
8. 不知道

12. (去年内，您)是否在门上或窗户上多加了锁？

0. 没有　　　　1. 是的
8. 不知道

13. (去年内，您)是否在家中安装了防盗警报器？

0. 没有　　　　1. 是的
8. 不知道

14. (去年内，您)是否接受了自我防范训练？

0. 没有　　　　1. 是的
8. 不知道

15. (去年内，您)是否在家中养了看门狗？

0. 没有　　　　1. 是的
8. 不知道

16. (去年内，您)是否购买了枪支或其他武器作为防卫？

0. 没有　　　　1. 是的
8. 不知道

17. 您不在家的时候，是否经常因为担心犯罪行为而开着灯，这样的情况在您家是？

1. 总是这样　　　　2. 经常这样
3. 几乎不这样　　　　4. 从来不这样
8. 不知道

18. 您不在家的时候，是否经常因为担心犯罪而要求您的邻居照看您的家，这样的情况在您家是？

1. 总是这样　　　　2. 经常这样
3. 几乎不这样　　　　4. 从来不这样

8. 不知道

19. 您是否经常因为担心犯罪而不去某个您本来想去的地方,这样的情况在您是?

1. 总是这样　　2. 经常这样

3. 几乎不这样　　4. 从来不这样

8. 不知道

20. 您是否经常因为担心犯罪避免独自外出,这样的情况在您是?

1. 总是这样　　2. 经常这样

3. 几乎不这样　　4. 从来不这样

8. 不知道

21. 您是否经常为了自卫而携带武器,这样的情况在您是?

1. 总是这样　　2. 经常这样

3. 几乎不这样　　4. 从来不这样

8. 不知道

警察、法院和监狱是犯罪司法系统的重要组成部分。您认为他们的工作做得怎么样?

22. 您认为警察的工作做得很好、好、一般还是差?

1. 很好[跳答第 23 题]　　2. 好[跳答第 23 题]

3. 一般　　4. 差

8. 不知道[跳答第 23 题]

22a. 您这样看的原因是因为____________________

23. 您认为法院的工作做得怎么样?(做得很好、好、一般还是差?)

1. 很好[跳答第 24 题]　　2. 好[跳答第 24 题]

3. 一般　　4. 差

8. 不知道[跳答第 24 题]

23a. 您这样看的原因是因为____________________

24. 监狱的工作做得怎么样?(做得很好、好、一般还是差?)

1. 很好[跳答第 25 题]　　2. 好[跳答第 25 题]

3. 一般　　4. 差

8. 不知道[跳答第 25 题]

24a. 您这样看的原因是因为____________________

25. 为了在马里兰州能建更多的监狱,您是否愿意每年多付 100 美金税?

0. 否　　1. 是　　8. 不知道

26. 并非每一个被判有罪的人都能被被捕入狱的。马里兰的监狱已经严重超员。一个在监狱的罪犯,每年的费用高达 17 000 美元。

正因为如此,许多有罪的人未被投入监狱,而是被判**每月**向假释官报告一次。

还有一种判法是,**每天**报告一次。

另一种判决法是,对这些人进行**全天**候的监控,随时掌握他们的行踪

请问,在这些判决法中,您认为哪一种最合适?

(年龄、可卡因的价值、任意的犯罪前科)

26a. 假如一个年龄为________岁的人，出售了价值________美元的可卡因，且犯有________次前科。您认为这个人应该被投入监狱、要求每月向假释官报告一次、每天报告一次，还是实行全天候监控，随时掌握他的行踪？

1. 投入监狱
2. 要求**每月**报告一次[跳答第 27a]
3. 要求**每天**报告一次[跳答第 27a]
4. **全天**候监控，随时掌握他的行踪[跳答第 27a]
8. 不知道

26w. 如果他在监狱里已经服了一段时间的刑，那么其余的刑期应该在狱外受全天候监控，还是应该在监狱继续服完全部刑期？

1. 出狱监控
2. 在监狱服完全部刑期
8. 不知道

武器也是任意的

27a. 假如一个年龄为________岁的人，在街头抢劫人某人的 100 美金。他还用某种武器________威胁受害人，且犯有________次前科。您认为这个人应该被投入监狱、要求每月向假释官报告一次、每天报告一次，还是实行全天候监控，随时掌握他的行踪？

1. 投入监狱
2. 要求每月报告一次[跳答第 28a]
3. 要求每天报告一次[跳答第 28a]
4. 全天候监控，随时掌握他的行踪[跳答第 28a]
8. 不知道

27w. 如果他在监狱里已经服了一段时间的刑，那么其余的刑期应该在狱外受**全天**候监控，还是应该在监狱继续服完全部刑期？

1. 出狱监控
2. 在监狱服完全部刑期
8. 不知道

财物价值是任意的

28a. 假如一个年龄为________岁的人犯了入室偷盗罪，偷盗了价值________美金的财物，且犯有________次前科。您认为这个人应该被投入监狱、要求**每月**向假释官报告一次、**每天**报告一次，还是实行**全天**候监控，随时掌握他的行踪？

1. 投入监狱
2. 要求每月报告一次[跳答第 29a]
3. 要求每天报告一次[跳答第 29a]
4. 全天候监控，随时掌握他的行踪[跳答第 29a]
8. 不知道

28w. 如果他在监狱里已经服了一段时间的刑，那么其余的刑期应该在狱外受**全天**候监控，还是应该在监狱继续服完全部刑期？

1. 出狱监控

2. 在监狱服完全部刑期

8. 不知道

伤害的不同严重程度

29a. 假如一个年龄为________岁的人犯有人身伤害罪。医生要求受害者(接受或无需接受)________治疗。罪犯有________次前科。

您认为这个人应该被投入监狱、要求**每月**向假释官报告一次、每天报告一次,还是实行全天候监控,随时掌握他的行踪?

1. 投入监狱

2. 要求**每月**报告一次[跳答第 29a]

3. 要求**每天**报告一次[跳答第 29a]

4. **全天**候监控,随时掌握他的行踪[跳答第 29a]

8. 不知道

29w. 如果他在监狱里已经服了一段时间的刑,那么其余的刑期应该在狱外受**全天**候监控,还是应该在监狱继续服完全部刑期?

1. 出狱监控

2. 在监狱服完全部刑期

8. 不知道

30. 您认为造成马里兰州街头暴力犯罪的主要原因是什么?

* * **只记录一条答案**

____

____

31. 除了建监狱之外,如果您可以向州长和立法机关提两条治理犯罪问题的建议,你的建议是什么?

____

____

D1. 最后我想问几个有关您个人背景的问题。

包括您自己在内,您家中共有几个年龄在 18 岁或 18 岁以上的人?

01 ~ 10　记录实际数目

11　10 人以上

99　非应用值—拒绝回答

D2. 您家中共有几个 18 岁以下的孩子?

0 ~ 7　记录实际数目

8　7 个以上

99　非应用值—拒绝回答

D3. 您生于哪一年?

00　1900 年以前

01 ~ 74　19 ____

88　不知道

99　非应用值—拒绝回答

D4. 您的最终学历是?

0　　无学历
1 ~ 7　　几年小学
8　　小学毕业
9 ~ 11　　几年中学
12　　中学毕业
13 ~ 15　　几年大学
16　　大学毕业
17　　几年研究生
18　　研究生毕业或专业技术学位
99　　拒绝回答

D5a. 你是否是西班牙人或是西班牙人的后裔?
0. 否　　1. 是

D5b. 你是
1. 白人　　2. 黑人
3. 亚洲人[如果 D5a 的答案是"是",不用问此题。]
4. 其他种族(具体说明):________
5. 拒绝回答

D6. 您的住房是自己的,还是租赁的?
1. 自己的　　2. 租赁的
3. 其他　　9. 拒绝回答

D7. 您现在的婚姻状况是:
1. 结婚　　2. 分居
3. 离婚　　4. 丧偶
5. 从未结过婚　　9. 非应用值—拒绝回答

D8. 您现在的工作是:
1. 全职　　2. 兼职
3. 无工作　　9. 非应用值—拒绝回答

D9. 您在目前的住址一共住了多少年?
00　　不到一年
01 ~ 50　　记录实际居住年数
51　　50 年以上
88　　不知道
99　　非应用值 - 拒绝回答

D10. 您家中共有多少个不同的电话号码是非营业性的?
1 ~ 6　　如实记录
7　　7 个以上
8　　不知道
9　　拒绝回答

D11. 1991 年,所有的家庭成员的年收入加在一起,您家中税前的年总收入在30 000 美金以上?

0. 否[跳答 D11a]　　1. 是[跳答 D11c]
9. 拒绝回答

D11a. 多于 20 000 美金
0. 否[跳答 D11b]　　1. 是[跳答 D12]
9. 拒绝回答

D11b. 多于 12 000 美金
0. 否[跳答 D12]　　1. 是[跳答 D12]
9. 拒绝回答

D11c. 多于 50 000 美金
0. 否[跳答 D12]　　1. 是[跳答 D11d]
9. 拒绝回答

D11d. 多于 75 000 美金
0. 否[跳答 D12]　　1. 是[跳答 D11e]
9. 拒绝回答

D11e. 多于 100 000 美金
0. 否[跳答 D12]　　1. 是[跳答 D12]
9. 拒绝回答

D12. 您住在哪一个县?

D13. 您的电话号码是[填入样本电话号码]
0. 否,那么我打的电话号码是?(确定具体号码)
1. 是　　9. 拒绝回答

我要问的问题就这么多。耽误了您很多时间,十分感谢。

美国舆论调查研究会职业道德和行业行为法规[1]

American Association for Public Opinion Research Code of Professional Ethics and Practices

附录 C

我们，美国舆论调查研究会的成员，同意下面的法规所阐述的原则。我们的目的是在进行舆论调查；将诸如这样的调查用于公共部门或私营部门的政策制定或决策，以及增进公众对调查方法的理解和恰当使用研究结果等方面，厉行合理而有道德的行业规范。

我们发誓，在进行分析和报告我们的工作时，在对待我们和被调查人、客户、那些最终会将我们的研究用于决策的目的的人们和一般的公众的关系时，我们将恪守科学和诚信的高标准。我们进一步承诺，我们将拒绝所有与本法规规定的原则不相一致的任务和指令。

法　规

Ⅰ. 专业工作的行为原则

A. 我们将认真负责地进行研究和调查工具的设计，并在数据收集、处理和分析中，采用所有合理步骤以确保结果的信度和效度。

1. 我们将推荐和使用那些根据我们的专业判断，被认为适合我们手头研究的问题的工具和方法。

2. 我们决不使用那些将得出错误的结论的研究工具和分析方法

3. 我们既不会有意对调查结果做出与得到的数据不相一致的解释，也不会对这样的解释缄默不语。

4. 我们决不有意超过数据实际所能保证的置信度，夸大结果的可信度。

B. 我们将在所有的研究报告中，依据部分Ⅲ所规定的最低要求，精确而详略适当地描述我们的方法和发现。

C. 如果我们工作的任何方面涉嫌违反本法规的有关规定，那么经 AAOR 执行委员会的批准，我们将提供有关调查的其他材料，其详细程度足以使调查业的同道对该调查做出专业的评价。

Ⅱ. 处理人际关系的职业职责原则

D. 公开：

1. 在准备公开我们的研究报告时，我们将确保我们的发现和解释或含义是对调查结果公允和精确的描绘。

2. 如果发觉公开的报告严重失实，即描述不正确、歪曲或不完整，我们将公开要求对不够精确之处进行修正，公开的、恰当的方式和场所包括：在公共传媒发表声明，向立法组织、行政机关或其他有关的团体，即向接受不精确报告的机构递交声明。

3. 我们将通知客户，根据 AAPOR 的规定，其成员，凡执行结果将公开发表的调查，必须公开有关诸如这样的调查最低限度的信息，因而我们将尽一切努力，说服他们同意执行这一规定，在他们公开发表调查结果时，最低限度地公开有关调查的信息。

E. 客户或主办者：

1. 在为一个私人客户工作时，除非在客户有明确指示和授权可公开散发时，或根据本法规部分 I 的 C 和 D 中的有关条款必须公开时，我们将对我们得到的有关客户和为该客户而做的操作和研究发现严格保密[2]。

2. 我们将注意我们采用的技术和自身能力的局限性，并将只接受我们的技术和能力所能及的研究任务。

F. 职业：

1. 我们认识到我们必须对民意研究的科学负责，尽可能坦率地散播源自我们的研究的思想和发现。

2. 本协会的成员资格不可用作专业能力的证明，因为协会并不在这方面为任何人或任何组织做出保证。

G. 被调查人：

1. 我们将避免那些可能伤害、侮辱或严重误导被调查人的行为和方法。

2. 我们将保证他们不受到任何人身伤害和/或不必要地和多余地涉及他们的隐私。

3. 除了少数例外，参与调查是自愿的。我们将向样本个人足够详细地介绍我们的调查，以允许他们在知情之后自由地做出是否参与的决定。

4. 我们将如实介绍我们的研究，不借研究之名，从事其他的活动（如推销、集资或政治竞选）。

5. 除非被调查人为特定的用途放弃保密，否则我们将把它作为被调查人的特权，并对答案中所有可能用于确定被调查人个人的信息严格保密。我们将不会因非研究的目的而暴露或使用被调查人的姓名，除非被调查人授权我们可以这样做。如果我们要求得到这样的许可，我们必须给被调查人以充分的理由。

6. 我们了解，将我们的调查结果用于合法的学报，并不能免除我们对所有可确认被调查人身份的信息保密的道德责任，或减少被调查人匿名问题的重要性。

Ⅲ. 必须公布的内容的最低要求

所有的研究报告，在报告发表的时候，内容中必须含包括有关调查研究自身和执行情况的基本信息。所有的民意调查研究者，都应遵守这一好的规范，并把它视

作自己的责任。我们应该公诸于众的信息，至少应该包括以下方面：

1. 调查的组织者和实施者。

2. 提出的问题的确切措辞，包括任何前述的说明、对调查员或被调查人进行的有理由预期可能对答案有所影响的解释。

3. 研究总体的定义和用以确定这一总体的抽样框。

4. 阐明抽样设计，清楚地说明研究者用以选取被调查人的方法。

5. 样本量，恰当的、合格的被调查人的标准和筛选方法，以及根据 AAPOR 的标准定义计算的回答率。最低限度，还应该提供样本部署的概括情况，以便能迅速地计算回答率。

6. 对发现的精确度进行必要的讨论，如果可以的话，应该包括抽样误差的估计值、使用的加权或估计的方法。

7. 哪些调查结果以部分样本，而非整个样本为依据，这些部分样本的大小。

8. 数据收集的方法、地点和时间。

AAPOR 理事会也许会就成果的发表、设计和调查的实施不定期地颁布一些指导方针和建议。

2004 年 3 月

注　释

1　美国舆论调查研究会（AAOR）曾于 2004 年秋天，就这一版本的法规投票表决。

2　州立大学和其他公共机构一般只能做在一个规定的时间之后，必须将结果公诸于众这样的调查。

术语表

偏倚(bias):一种恒高或恒低于真实总体的测量或估计偏向。

回拨验证(callback verification):一种质量控制的手段。从每一调查员已经完成的个案抽取一个样本进行回访,以验证调查是否如实进行。

普查(census):一种在总体的所有单位或成员中收集信息的方法。

封闭式问题(closed-ended questions):被调查人对问卷题项的回答都选自一张特定的答案清单的问题。

聚类、丛、群(cluster):单元的自然聚集,如街区中的住户。在抽样框中它与若干个总体成分只选择其中一个进入样本的问题有关——例如列举代表个体丛的学生丛或住户的系列清单。

编码(coding):将数字分配给被调查人对调查问题给出的答案。

认知访谈(cognitive interview):一种一对一式的访谈,被设计用于确定被调查人在回答问题时使用的方法步骤,或确定被调查人在理解或回答一个问题时发生的问题。这样的访谈使用各种实验室技术,如边想边说思维法(think aloud procedures)或问题解释(the paraphrasing of questions)等。

置信区间(confidence interval):某一包含总体值,有特定概率的高于或低于样本估计值的值域,

合作率(cooperation rate):相对于全部完成、部分完成和拒访的被调查人之和的全部完成的调查数。

数据检查(cata checking):一种控制调查质量的手段。在全部调查完成之前,通过阅读调查数据、检查频数分布,或有选择地的制作跨变量的交互分类表,探测数据中存在的错误或异常。

数据简编(data reduction):为进一步分析而对数据文件进行的编码和各种准备。

因变量(dependent variable):目标总体的一种为研究所要估计或解释的,作为自变量的一种函数的特征或特性。

设计效应(design effect):一种用于比较在样本容量相同时,特定的抽样设计和简单随机抽样(srs)的精确统计量。它是特定的样本(分子)与相同样本容量的简单随机样本(分母)的方差的比率。

双筒(双管)式问题(double -barreled question):一种提出两个问题的问卷题项形式,被调查人只需要对其中的一个给出答案。

估计计划(estimation plan):用来自样本调查的数据计算整个总体或子群体的方法。

焦点组(focus group):一组具有相同特点的人被召集在一起,在某一论题或问题上进行有导向性的讨论。焦点团体经常被作为问卷设计过程中的一个程序。

看门人(gatekeeper):在电话调查中那位接电话和控制与指定的被调查人联系的人。控制的方法是不允许调查员和选定的被调查人通话。在面对面的调查中也可能有看门人存在。

监测(monitoring):一种质量控制的方法。监测人员对正在进行的电话调查进行监听。检查调查员是否按规定的要求进逆行调查。它也可用于试调查阶段,作为问卷设计的一种工具,确定问卷中存在的问题。

多次性(multiplicity):一个调查总体中的某些成员多次出现在抽样框中。

综合性调查(omnibus survey):一种有多重目的的研究。目的在于对一组问题进行探究,和/或收集涉及领域广泛的、有关多个题目,且题目之间互不相关的信息的调查。

开放式问题(open-ended questions):问卷中那些由被调查人自行决定答案的文字的问题。

操作化(operationalization):定义一个概念,如社会经济地位,并确定它在调查中如何进行测量。如果将它定义为教育和收入,那么操作化的概念在不仅包括问及被调查人这方面信息的问题,同时还应包括如何将这些信息融合的方法。

总体超额(overcoverage):抽样清单中包括了那些并不属于定义的调查总体的成员的实体。

过度报告(overreporting):被调查人的陈述言过其实,例如他们报告的收入高于实际,或某些实际活动或实际行为多于实际情况。

参与式调查(participating pretest):一种被调查人在调查之前,被告知它是调查的一个组成部分的试调查。

概率水平(置信水平)[probability level(confidence level)]:样本估计值落入置信区间规定的值域的可能性。

选择的概率(probability of selection):总体的每一单元或元素被选入样本的可能性。其含义等同于选择的机会和包含的概率。

概率抽样(probability sampling):一种给每一成员以一个已知(非零)的包括的机会和采用随机的选择机制抽样方法。

探查性问题(probe a question):一种设计用于引出一个已经提供了可能不甚完整或不够清楚的答案,或无法归入规定的答案类别的答案的回答人的答案的问题。

抄本、协议(protocol):由边想边说思维(认知访谈)阶段的录音转录的文字。诸如这样的计划或程序应该安排在一个调查或对一个被调查人调查之后。这一名词有时也会有其他不同的含义,特指执行认知访谈的一系列程序、指示或探询等。

随机数码拨号(random digit dialing,RDD):一种用随机方法为电话调查选取或生成电话号码最后 1 ~4 位数字的方法。

回答偏倚(response bias):估计的真值和实际回答的期望值之间的差异。

回答效应(response effect):在回答一个问题时与某一特殊因素有关的误差量。诸如这样的因素包括缺失的记忆或自我表现等。

回答率(response rate):完成问卷的样本成员数除以合格的样本成员总数的百分比。

抽样误差(sampling error):存在于总体参数的样本估计值中的误差。其产生原因仅仅是因为估计依据的测量值取自样本,而非整个目标总体。

抽样框偏倚(sampling frame bias):因为抽样框列举的内容非随机性而导致的一种样本偏倚。

抽样比率(sampling rate):为达到要求的样本容量而必须从抽样框中选取的数目除以抽样框选取总数,通常以分数形式出现。

筛选性问题(screening questions):用于确定成员人数和符合调查目标总体的成员的问题,或用于确定是否应该向被调查人询问问卷中的某一组问题的问题。

标准差(standard deviation):样本方差的平方根。

分层抽样(stratified sampling):一种抽样方法。样本选自一个已经根据将要测量或估计的研究变量分组或分层的总体。就关系这些变量的某些特征而言,组或层内的个体具有比较高的同质性。

子群体(subgroups):一个来自比较大的总体的个体的子组。组内的个体具有相同的特征,譬如女性、黑人或年龄在65岁以上。

调查误差(survey error):样本估计值偏倚和方差的联合效应,通常称为调查误差。

边想边说(think aloud):一种调查者和被调查人一对一相处的方式。这种方式要求被调查人在回答问题时,一边回忆一边回答;多用于问卷的设计和改进阶段。参见认知访谈条。

未宣式试调查(undeclared pretest):一种被调查人不自觉参与的设计改进调查方案的试调查。

总体差额(undercoverage):因某些已定义的总体成员未包含在抽样框内而产生的差额,也可能由于数据收集的方法或调查的程序不当而引起。

低报(underreporting):被调查人报告情况低于实际情况,如报告的某些事物少于他们实际所有,或报告的某些活动或行为常常低于他们实际所为。

方差(variance):反映样本或总体散布或差异的程度;其差异出现在同一程序的重复试验中。

加权法(weighting procedure):一种用人口学的子群体或其他与无偏估计法有关的因素,对样本成员选择的不等、调查完成率的不等进行修正的方法。

参考文献

Aday , L. A. (1996). *Designing and conducting health surveys: A comprehensive guide.* San Francisco: Jossey-Bass.

American Association for Public Opinion Research (AAPOR). (1998). *Standard definitions: Final dispositions of case codes and outcome rates for RDD telephone surveys and in-person household surveys.* Ann Arbor, MI: AAPOR.

American Association for Public Opinion Research (AAPOR). (2004). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (3rd ed.). Lenexa, KS: AAPOR.

Aneshensel, C. S., Frerichs, R. R., Clark, V. A., & Yokopenic, P. A. (1982). Measuring depression in the community: A comparison of telephone and personal interviews. *Public Opinion Quarterly*, 46, 110-121.

Belinfante, A. (2004). *Telephone penetration by income by state.* (FCC News Release). Washington, DC: Industry Analysis and Technology Division, Wireline Competition Bureau, Federal Communication Commission.

Belson, W. A. (1981). *The design and understanding of survey questions.* Brookfield, VT: Gower.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (1991). *Measurement errors in surveys.* New York: John Wiley & Sons.

Biemer, P. P., Lyberg, L. E. (2003). *Introduction to survey quality.* New York: John Wiley & Sons.

Binson, D., Murphy, P. A,, & Keer, D. (1987, May). *Threatening questions for the public in a survey about AIDS.* Paper presented at annual meeting of American Association for Public Opinion Research, Hershey, PA.

Bischoping, K. (1989). An evaluation of interviewer debriefing in survey pretests. In C. Cannell, L. Oksenberg, F. Fowler, G. Kalton, &: K. Bischoping (Eds.), *New techniques for pretesting survey questions.* (Final Report for Grant HS05616, National Center for Health Services Research). Ann Arbor: University of Michigan, unpublished report.

Blair, J., & Chun, Y. (1992). *Quality of data from converted refusals in telephone surveys*. Paper presented at the American Association for Public Opinion Research Conference, St. Petersburg, FL.

Blair, J., Menon, G., & Bickart, B. (1991). Measurement effects in self vs. proxy response to survey questions: An information-processing perspective. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 145-166). New York: John Wiley & Sons.

Bradburn, N. M., & Miles, C. (1979). Vague quantifiers. *Public Opinion Quarterly*, 43(1), 92-101.

Bradburn, N. M., Sudman, S., & Associates. (1979). *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. San Francisco: Jossey-Bass.

Brick, J. M., Waksberg, J., Kulp, D., & Starer, A. (1995). Bias in list-assisted telephone samples. *Public Opinion Quarterly*, 59(2), 218-235.

Bryant, B. E. (1975). Respondent selection in a time of changing household com position. *Journal of Marketing Research*, 12, 129-135.

Campanelli, P. C., Rothgeb, J. M., Esposito, J. L., & Polivka, A. E. (1991). *Methodologies for evaluating survey questions: An illustration from a CPS CATI/RDD Test*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, AZ.

Campbell, B. A. (1981). Race-of-interviewer effects among southern adolescents. *Public Opinion Quarterly*, 45, 231-244.

Cannell, C. F., Fowler, P. J., & Marquis, K. H. (1968). The influence of interviewer and respondent psychological and behavioral variables on the reporting in house hold interviews. *Vital and Health Statistics*, Series 2, No. 26. Washington, DC: U. S. Government Printing Office.

Cannell, C., Oksenberg, L., Fowler, F., Kalton, G., & Bischoping, K. (1989). *New techniques for pretesting survey questions*. (Final Report for Grant HS05616, National Center for Health Services Research). Ann Arbor: University of Michigan, Unpublished report.

Caron, J. (1992). *An introduction to psycholinguistics* (T. Pownall, Trans.). New York: Harvester Wheatsheaf. (Original work published 1989)

Carroll, R. (2003, August 4). Millions getting rid of landline phones. *Associated Press*.

Casady, R., & Lepkowski, J. (1993). Stratified telephone survey designs. *Survey*

Methodology, 19(1), 103-113.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Interviewer and question effects on sex items. *Public Opinion Quarterly*, 60(3), 345-375.

Cho, H., & LaRose, R. (1999). Privacy issues in Internet surveys. *Social Science Computer Review*, 17, 421-434.

Church, A. H. (1993). Incentives in mail surveys: A meta-analysis. *Public Opinion Quarterly*, 57(1), 62-79.

Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax and Web-based survey methods. *International Journal of Market Research*, 43(4), 441-452.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley & Sons.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1.

Cohen, J. (1988). *Statistical power analysis for the behav, ioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Conrad, F., & Blair J. (2003). Aspects of data quality in cognitive interviews: The case of verbal reports. In Presser, S., Rothgeb., J., Couper, M., Lessler, J., Martin, E., Martin, J., et al. (Eds.). *Methods for testing and evaluating survey questionnaires*. New York: John Wiley & Sons. 2004.

Conrad, F., & Blair, J. (2004). Data quality in cognitive interviews: The case of verbalreports. In Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., et al., (Eds.), *Methods for testing and evaluating survey questionnaires.* New York: John Wiley & Sons.

Conrad, F., Blair J., & Tracy E. (1999). *Verbal reports are data! A theoretical approach to the analysis of cognitive interviews*. Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference, Statistical Policy, Working Paper #30, Washington, DC, June 2000.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage.

Council of American Survey Research Organizations (CASRO). (1982). *Report of the CASRO completion rates task force*. New York: Audits and Surveys, Inc., Unpublished report.

Couper, M. P. (1997). Survey introductions and data quality. *Public Opinion-Quarterly*, 61(2), 317-338.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *PublicOpinion Quarterly*, 64(4), 464-494.

Couper, M. P. (2001). *Designing effective Web surveys.* Workshop presented at annual meeting of American Association for Public Opinion Research. Montreal, Quebec, Canada.

Couper, M. P., Baker, R. P., Bethlehem, J. Clark, C. Z. F., Martin, J., Nicholls, W. L., et al. (Eds). (1998). *Computer assisted survey information collection.* New York: John Wiley & Sons.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65(2), 230-253.

Current Population Survey. (2002). *Telephone in the household by poverty and income level by geographic region.* [MRDF] March Supplement 2002. Retrieved February 16, 2004 from http://ferret. bls. census. gov.

Curtin, R., Presser S., & Singer E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, 64 (4), 413-428.

Czaja, R. F. (1987-1988). Asking sensitive behavioral questions in telephone interviews. *International Quarterly of Community Health Education*, 8, 23-31.

Czaja, R. F., Blair, J., & Sehestik, J. P. (1982). Respondent selection in a telephone survey: A comparison of three techniques. *Journal of Marketing Research*, 21, 381-385.

Davis, D. K., Blair, J., Gourdreau, N., Boone, M. S., Johnson, L. R., & Robles, E. (1998). *Research on race and Hispanic origin for Census* 2000. Report to Population Division, Bureau of the Census.

Davis, J. A., & Smith, T. W. (1991). *General social surveys*, 1972-1991: *Cumulative codebook.* Chicago: National Opinion Research Center.

Davis, J. A., & Smith, T. W. (1993). *General social surveys*, 1972-1993: *Cumulative codebook.* Chicago: National Opinion Research Center.

Davis, J. A., Smith, T. W., & Marsden, P. V. (2001). *General social surveys codebook.* Principal investigator, James A. Davis; director and co-principal investigator, Tom W. Smith. Chicago: National Opinion Research Center.

de Leeuw, E. (1992). *Data quality in mail, telephone and face-to-face surveys.* Amsterdam: TT-Publikaties.

Demaio, T. J. (1983). *Approaches to developing questions.* (Statistical Policy Working Paper 10). Washington, DC: Statistical Policy Office, U.S. Office of Management and Budget.

Dillman, D. A. (1978). *Mail and telephone surveys: Tbe total design method.* New York: John Wiley & Sons.

Dillman, D. A. (2000). *Mail and Internet surveys. The tailored design method.* New York: John Wiley & Sons.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data, revised edition.* Cambridge, MA: The MIT Press.

Everitt, B., & Hay, D. (1992). *Talking about statistics: A psychologist's guide to design & analysis.* New York: Halsted Press.

Ezzati-Rice, T. M., White, A. A., Mosher, W. D., & Sanchez, M. E. (1995). *Time, dollars, and data: Succeeding with remuneration in health surveys.* Paper presented at the Federal Committee on Statistical Methodology Conference: Seminar on New Directions in Statistical Methodology, Washington DC.

Federal Communication Commission. (2003). Trends in telephone service. *In Telephone penetration and telephone subscribersbip* (pp 16-17). Washington, DC: Author.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley & Sons.

Fowler, F. J. (1984). *Survey research methods.* Newbury Park, CA: Sage.

Fowler, F. J. (1989). Coding behavior in pretests to identify unclear questions. In F. J. Fowler (Ed.), *Conference proceedings, health survey research methods* (pp. 9-12). Washington, DC: National Center for Health Services Research.

Fowler, F. J. (1995). *Improving survey questions: Design and evaluation.* Newbury Park, CA: Sage.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing.* Newbury Park, CA: Sage.

Frey, J. H. (1989). *Survey research by telephone.* Newbury Park, CA: Sage.

Glenberg, A. M. (1996). *Learning from data: An introduction to statistical reasoning* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Greenbaum, T. L. (1998). *The handbook for focus group research* (2nd ed.). Thousand Oaks, CA: Sage.

Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly*, 51, 5156-5172.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: John Wiley & Sons.

Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, II, W. L.,

& Waksberg, J. (Eds.). (1988). *Telephone survey methodology*. New York: John Wiley & Sons.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992) Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56(4), 475-495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. Somerset, NJ: John Wiley & Sons.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.). (2002). *Survey nonresponse*. New York: John Wiley & Sons.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (Eds.). (2004). *Survey methodology*. New York: John Wiley & Sons.

Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. Orlando, FL: Academic Press.

Guensel, P. J., Berckmans, T. R,, & Cannell, C. F. (1983). *General interviewing techniques*. Ann Arbor, Ml: Survey Research Center.

Hatchett, S., & Schuman, H. (1975-1976). White respondents and race-of-interviewer effects. *Public Opinion Quarterly*, 39, 523-528.

Hidiroglou, M. A., Drew, J. D., & Gray, G. B. (1993). A framework for measuring and reducing nonresponse in surveys. *Survey Methodology*, 19(1), 81-94.

Hippler, H. J., Schwarz, N., & Sudman, S. (1987). *Social information processing and survey methodology*. New York: Springer-Verlag.

Hochstim, J. R. (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association*, 62, 976-989.

Huggins, V., & Eyermen, J. (2001). *Probability based Internet surveys: A synopsis of early methods and survey research results*. Paper presented at the Federal Committee on Statistical Methodology Conference, Washington, DC.

Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines: Report of the advanced research seminar on cognitive aspects of survey methodology. Washington*, DC: National Academy Press.

James, J. M., & Bolstein, R. (1990). Monetary incentives and follow-up mailings. *Public Opinion Quarterly*, 54(3), 346-361.

Johnson, T. P., Parker, V., & Clements, C. (2001). Detection and prevention of data falsification in survey research. *Survey Research*, 32(3), 1-2.

Kalton, G. (1981). *Compensating for missing survey data.* (Report for Department of Health and Human Services Contract HEW-100-79-0127, Survey Research Center, Institute for Social Research). Ann Arbor: The University of Michigan.

Kalton, G. (1983). *Introduction to survey sampling.* Newbury Park, CA: Sage.

Kanninen, B. J., Chapman, D. J., & Hanemann, W. M. (1992). Survey data collection: Detecting and correcting for biases in responses to mail and telephone surveys. In *Conference Proceedings, Census Bureau* 1992 *Annual Research Conference* (pp. 507-522). Arlington, VA: U.S. Department of Commerce.

Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M. P. (Eds.). (1989). *Panel surveys.* New York: John Wiley & Sons.

Keeter, S., Miller C., Kohut A., Groves R. M., & Presser S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, 64(2), 125-148.

Kinsey, S. H., & Jewell, D. M. (1998). A systematic approach to instrument development in CAI. In M. Couper, R. Baker, J. Bethlehem, C. Clark, J. Martin, W. Nicholis, et al. (Eds.). *Computer assisted survey information collection* (pp. 105-123). New York: John Wiley & Sons.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Kish, L. (1987). *Statistical design for research.* New York: John Wiley & Sons.

Kulka, R. A. (1995). *The use of incentives to survey "hard-to-reach" respondents: A brief review of empirical research and current research practice.* Paper presented at the Federal Committee on Statistical Methodology Conference: Seminar on New Directions in Statistical Methodology, Washington DC.

Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials*, 2, 93-113.

Lachin, J. M. (1998). Sample size determination. In P. Armitage & T. Colton (Eds.), *Encyclopedia of biostatistics* (Vol. 5, pp. 3892-3903). New York: John Wiley & Sons.

Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys.* New York: John Wiley & Sons.

Locander, W., Sudman, S., & Bradburn, N. M. (1976). An investigation of interview method, threat, and response distortion. *Journal of the American Statistical Association*, 71, 269-75.

Lyberg, L., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., & Trewin, D.

(Eds.). (1997). *Survey measurement and process quality*. Somerset, NJ: John Wiley & Sons.

Mangione, T. W. (1995). *Mail surveys: Improving the quality*. Thousand Oaks, CA: Sage Publications.

Marsden, P. V. (Ed.). (1994). *Sociological methodology* (Vol. 24). Cambridge, MA: Blackwell.

Moser, C. A.,&: Kalton, G. (1972). *Survey methods in social investigation* (2nd ed.). New York: Basic Books.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-606.

Nicholls II, W. L., Baker, R. P., & Martin, J. (1997). The effect of new data collec tion technologies on survey data quality. In L. Lyberg, P. Biemer, M, Collins, E. de Leeuw, C. Dippo, N. Schwarz, et al. (Eds.), *Survey measurement and process quality* (pp. 221-248). New York: John Wiley & Sons.

Oldendick, R. W., Bishop, G. F., Sorenson, S. B., & Tuchfarber, A. J. (1988). A comparison of the Kish and last-birthday methods of respondent selection in telephone surveys. *Journal of Official Statistics*, 4(4), 307-318.

Oldendick, R. W., & Link, M. W. (1994). The answering machine generation: Who are they and what problem do they pose for survey research? *Public Opinion Quarterly*, 58(2), 264-273.

O'Rourke, D., & Blair, J. (1983). Improving random respondent selection in telephone surveys. *Journal of Marketing Research*, 20, 428-432.

Parsons, J., Owens, L., & Skogan,. (2002). Using advance letters in RDD surveys: Results of two experiments. *Survey Research*, 33(1), 1-2.

Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.

Piekarski, L. (1997). Unlisted telephone rates in 320 metropolitan statistical areas. *The Frame*. Survey Sampling International.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? *Sociological Methodology*, 24, 73-104.

Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., et al. (Eds.). (2004). *Methods for testing and evaluating survey questionnaires*. New York: John Wiley & Sons.

Rasinski, K. A., Mingay, D., &: Bradburn, N. M. (1994). Do respondents really mark all that apply? *Public Opinion Quarterly*, 58(3), 400-408.

Rizzo, L., Brick, J. M., & Park, I. (2004). A minimally intrusive method for sampling persons in random digit dial surveys. *Public Opinion Quarterly*, 68 (2), 267-274.

Rosner, B. A. (2000). *Fundamentals of biostatistics* (5th ed.). Pacific Grove, CA: Duxbury.

Royston, P. N. (1989). Using intensive interviews to evaluate questions. In F. J. Fowler (Ed.), *Conference proceedings, health survey research methods* (pp. 3-8). Washington, DC: National Center for Health Services Research.

Salmon, C. T., &: Nichols, J. S. (1983). The next-birthday method of respondent selection. *Public Opinion Quarterly*, 47, 270-276.

Sarndal, C. -E., Swensson, B., & Wretman, J. (1992). *Model-assisted survey sampling*. New York: Springer-Verlag.

Schaefer, D., & Dillman, D. A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly*, 62 (3), 378-397.

Schechter, S., & Blair, J. (2001). Expanding cognitive laboratory methods to test self-administered questionnaires. *International Journal of Cognitive Technology*, 6(2), 26-32.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Orlando, FL: Academic Press.

Schwarz, N., Strack, F., & Mai, H. -P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational analysis. *Public Opinion Quarterly*, 55(1), 3-23.

Schwarz, N., & Sudman, S. (Eds.). (1992). *Context effects in social and psychological research*. New York: Springer-Verlag.

Schwarz, N., & Sudman, S. (Eds.). (1994). *Autobiographical memory and the validity of retrospective reports*. New York: Springer-Verlag.

Schwarz, N., & Sudman, S. (Eds.). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.

Schwarz, N., & Wellens, T. (1997). Cognitive dynamics of proxy reporting: The diverging perspectives of actors and observers. *Journal of Official Statistics*, 13 (2), 159-179.

Shettle, C., & Mooney, G. (1999). Monetary incentives in U. S. government surveys. *Journal of Official Statistics*, 15(2), 231-250.

Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, 64(2), 171-188.

Sirken, M. G., Herrmann, D. J., Schechter, S., Schwarz, N., Tanur, J. M., & Tourangeau, R. (Eds.). (1999). *Cognition and survey research*. New York: John Wiley & Sons.

Strack, F. (1992). "Order effects" in survey research: Activation and information functions of preceding questions. In N. Schwarz & S. Sudman (Eds.). *Context effects in social and psychological research* (pp. 23-34). New York: Springer-Verlag.

Sudman, S. (1967). *Reducing the cost of surveys*. Chicago: Aldine.

Sudman, S. (1976). *Applied sampling*. New York: Academic Press.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Tanur, J. M. (1992). *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation.

Tourangeau, R., Couper, M. P., & Steiger, D. M. (2001). *Social presence in web surveys*. Paper presented at the Federal Committee on Statistical Methodology Conference, Washington, DC.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299-314.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format and question context. *Public Opinion Quarterly*, 60(2), 275-304.

Tripplett, T. (1994, May). *What is gained from additional call attempts and refusal conversion and what are the cost implications?* Paper presented at the International Field Directors and Field Technicians Conference, Boston, MA.

Troldahl, V. C., & Carter, Jr., R. E. (1964). Random Selection of Respondents within Households in Phone Surveys. *Journal of Marketing Research*, 1(2), 71-76.

Tucker, C., Lepkowski, J. M., & Piekarski, L. (2002). The current efficiency of list assisted telephone sampling designs. *Public Opinion Quarterly*, 66(3),

321-338.

Turner, C. F., & Martin E. (Eds.). (1984). *Surveying subjective phenomena* (Vol. 1). New York: Russell Sage Foundation.

U. S. Census Bureau. (1989). *Bicentennial census facts*: 1990 *Decennial census*. Washington, DC: U. S. Government Printing Office.

U. S. Census Bureau. (1993). 1990 *Census of populati on and housing*: *Population and housing unit counts—North Carolina* (1990 CPH-2-35). Washington, DC: U. S. Government Printing Office.

U. S. Census Bureau. (2000a). *Profile of selected housing characteristics*: 2000 [MRDF] Quick tables—American FactFinder. Retrieved February 6, 2004 from http://factfinder. census. gov.

U. S. Census Bureau. (2000b). *Telephone service available*. [MRDF] Detail tables— American FactFinder. Retrieved February 6, 2004 from http://factfinder. census. gov.

U. S. Census Bureau. (2000c). *Tenure by poverty status in* 1999 *by telephone service available*. [MRDF] Detail tables—American FactFinder. Retrieved April 29, 2004 from http://factfinder. census. gov.

U. S. Census Bureau. (2000d). *Profile of general demographic characteristics*: 2000. [MRDF] Quick Tables-American FactFinder. Retrieved February 9, 2004 from http://factfinder. census. gov.

U. S. Census Bureau. (2003). *Annual demographic survey*: *Current population survey March supplement* 2002. Washington, DC: U. S. Dept. of Commerce (http://www. bls. ccensus. gov/cps/datamain. htm).

U. S. Census Bureau. (2004). *Meeting* 21*st century demographic data needs—Implementing the American community survey. Report* 4: *Comparing general demographic and housing characteristics with census* 2000. Washington. DC: U. S. Government Printing Office.

Waksberg, J. (1978). Sampling methods for random digit dialing. *Journal of the American Statistical Association*, 73(361), 40-46.

Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). Cash versus lotteries versus charities in mail surveys. *Public Opinion Quarterly*, 60(4), 542-562.

Willimack, D., Lyberg, L., Martin, J., Japec, L., & Whitridge, P. (2004). Evolution and adaptation of questionnaire development: evaluation and testing methods in establishment surveys. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al. (Eds.). *Methods for testing*

and evaluating survey questionnaires. New York: John Wiley & Sons.

Willis, G. (1994). *Cognitive interviewing and questionnaire design: A training manuall.* Working paper series. Hyattsville, MD: National Center for Health Statistics.

Willis, G. B., Trunzo, D. B., & Strussman, B. J. (1992). The use of novel pretesting techniques in the development of survey questionnaires. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 824-828.

Wiseman, F. (1972). Methodological bias in public opinion surveys. *Public Opinion Quarterly*, 36(1), 105-108.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). A meta-analysis of mail surveys. *Public Opinion Quarterly*, 55(4), 613-639.

［Ｇｅｎｅｒａｌ Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝抽样调查设计导论
作者＝罗纳德·扎加　约翰尼·布莱尔著　沈崇麟译
页数＝２５５
ＳＳ号＝１１８６８１９５
出版日期＝２００７年０４月第１版